

Gencore version 4.5							
Copyright (c) 1993 - 2000 Compugen Ltd.							
<b>OM nucleic - nucleic search, using sw model</b>							
<b>Run on:</b> September 15, 2002, 21:32:11 ; search time 2741.47 Seconds							
(without alignments) (61617.753 Million cell updates/sec)							
<b>Title:</b> US-09-828-313-22	2177	100	0	2177	6	AZ280388	AZ280978 Sequence
<b>Perfect score:</b> 2177	24	8	683	6	AZ280375	AZ29315 Homo sapi	
<b>Sequence:</b> 1 atcccggttggatggct.....gcagatattgcggtaacgc 2177	2	1	12520	9	AC024967	AC024964 Homo sapi	
<b>Scoring table:</b> IDENTITY_NUC	42.8	2.0	157442	2	AC024964	AC024964 Homo sapi	
<b>Searched:</b> 1797056 seqs, 10463268293 residues	42.8	2.0	183978	2	AC024967	AC024967 Homo sapi	
<b>Total number of hits satisfying chosen parameters:</b> 3595312	42.8	2.0	70203	2	AC024967	AC024967 Homo sapi	
<b>Post-processing:</b> Maximum Match 0%	42.8	1.8	153072	9	AC024967	AC024967 Homo sapi	
<b>Listing first 45 summaries</b>							
<b>Minimum DB seq length:</b> 0	10	40	1.8 14200	2	AC02297	AC02297 Homo sapi	
<b>Maximum DB seq length:</b> 200000000	10	40	1.8 125020	9	AF29315	AF29315 Homo sapi	
<b>Database :</b> GenBank	11	38	1.8 15542	30	AC013612	AC013612 Homo sapi	
1: qb_ba:*	12	38	1.8 15633	2	AC07067	AC07067 Homo sapi	
2: qb_htg:*	13	38	1.8 15633	9	AC00813	AC00813 Homo sapi	
3: qb_in:*	14	38	1.8 163914	9	AC00813	AC00813 Homo sapi	
4: qb_om:*	15	37	1.7 69507	2	AC079266	AC079266 Mus muscu	
5: qb_ov:*	16	37	1.7 73465	9	HS256M13	HS256M13 Homo sapi	
6: qb_pat:*	17	37	1.7 16182	2	HS214G14	HS214G14 Homo sapi	
7: qb_ph:*	18	37	1.7 34000	9	HS21C004	HS21C004 Homo sapi	
8: qb_pl:*	19	37	1.7 6706	9	AL9311	AL9311 Sequence 42	
9: qb_pr:*	20	37	1.7 91940	9	AL445674	AL445674 Human DNA	
10: qb_ro:*	21	37	1.7 157611	2	AC019204	AC019204 Homo sapi	
11: qb_stc:*	22	37	1.7 15307	10	AL592112	AL592112 Mouse DNA	
12: qb_sy:*	23	37	1.7 171726	9	AC007052	AC007052 Homo sapi	
13: qb_un:*	24	37	1.7 181750	2	AC025506	AC025506 Homo sapi	
14: qb_vl:*	25	37	1.7 193342	2	AC027588	AC027588 Homo sapi	
15: qb_ba:*	26	37	1.7 196869	2	AC087535	AC087535 Homo sapi	
16: qb_fun:*	27	37	1.7 20774	2	AP001592	AP001592 Homo sapi	
17: qb_hum:*	28	37	1.7 155163	2	AC058990	AC058990 Rattus no	
18: qb_in:*	29	37	1.7 15825	2	AC023503	AC023503 Homo sapi	
19: qb_mu:*	30	37	1.7 164481	2	AC034152	AC034152 Homo sapi	
20: qb_om:*	31	37	1.7 173794	9	AC023108	AC023108 Homo sapi	
21: qb_or:*	32	37	1.7 175968	9	AL161905	AL161905 Human DNA	
22: qb_cv:*	33	36	1.7 133370	9	AC055707	AC055707 Homo sapi	
23: qb_pat:*	34	36	1.7 152318	9	AL162851	AL162851 Homo sapi	
24: qb_ph:*	35	36	1.7 169141	9	AC021538	AC021538 Homo sapi	
25: qb_pl:*	36	36	1.7 171911	2	AC025637	AC025637 Homo sapi	
26: qb_ro:*	37	36	1.7 172282	9	AC010304	AC010304 Homo sapi	
27: qb_sts:*	38	36	1.7 163167	9	AC011347	AC011347 Homo sapi	
28: qb_un:*	39	36	1.7 186516	2	AC096938	AC096938 Rattus no	
29: qb_vl:*	40	36	1.7 182882	9	AF207069	AF207069 Homo sapi	
30: qb_sy:*	41	36	1.7 119800	9	AP001482	AP001482 Homo sapi	
31: qb_htg_inv:*	42	36	1.7 126662	9	AC004876	AC004876 Homo sapi	
32: qb_htg_other:*	43	36	1.7 167191	2	AC093442	AC093442 Papio cyn	
33: qb_htg_inv:*	44	36	1.7 168187	9	AL1139092	AL1139092 Human DNA	
	45	36	1.7 189058	2	AC013339	AC013339 Homo sapi	
<b>ALIGNMENTS</b>							
<b>RESULT</b>	1						
AX280988	AX280988	AX280988	AX280988	AX280988	AX280988	AX280988	AX280988 Sequence
DEFINITION	Sequence	22	from Patent	W0177356.	2177 bp	DNA	linear
ACCESSION	AX280988	AX280988	AX280988	AX280988	GI:16608262	PAT	02-NOV-2000
VERSION	AX280988.1						
KEYWORDS	.						
SOURCE	Physcomitrella Patens						
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Funariales; Funariaceae; Physcomitrella						
REFERENCE	1 (sites)						
AUTHORS	da costa e Silva,O., Bohnert,H.J., van Thiel,N., Chen,R. and Sarría-Millán,R.						
TITLE	Protein kinase stress-related Proteins and methods of use in plant						
JOURNAL	Patent: WO 017356 A 22 18-OCT-2001;						
FEATURES	Location/Qualifiers						
Source	1. -2177 "/organism="Physcomitrella patens" "/db_xtref="taxon:3218"						
BASE COUNT	522 a	508 c	547 g	600 t			
ORIGIN							
Result No.	Score	Query Length	DB ID	Description			
				SUMMARIES			







JOURNAL						
COMMENT						
TITLE						
ROY,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Marquis,N., Levin,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margolis,N., McCarthy,M., McEwan,J., McGurk,A., McKernan,K., McPheefer,R., McRae,T., Meneus,L., Mihowa,T., Miranda,C., Mileng,V., Morrow,J., Murphy,T., Navljar,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peltier,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testafe,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassilev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A., and Zody,M.						
RESEARCH	Submitted 03-MAR-2000; Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA					
ON MAY 7, 2001 this sequence version replaced g1:9954795.	All repeats were identified using RepeatMasker:					
SMIT, A.F.A. & Green, P. (1996-1997)	http://ftp.genome.washington.edu/RM/RepeatMasker.html					
Center project name: L6081	Center clone name: 554_B_18					
Center code: WIBHEAD	Institute/ MIT Center for Genome Research					
Web site: http://www-seq.wi.mit.edu						
Contact: sequence_submissions@genome.wi.mit.edu	Project Information					
	----- Summary Statistics -----					
	Sequencing vector: M13; M77815; 100% of reads					
	Chemistry: Dye-terminator	BIG Dye;	100% of reads			
	Assembly program: phrap; version 0.96031					
	Consensus quality: 17705 bases at least Q40					
	Consensus quality: 180493 bases at least Q30					
	Consensus quality: 181650 bases at least Q20					
	Insert size: 182000; agarose-fp					
	Insert size: 182378; sum-of-contigs					
	Quality coverage: 5.1 in Q20 bases; sum-of-contigs					
	Quality coverage: 5.1 in Q20 bases; sum-of-contigs					
	-----					
	* NOTE: This is a 'working draft' sequence. It currently					
	* consists of 17 contigs. The true order of the pieces					
	* is not known and their order in this sequence record is					
	* arbitrary. Gaps between the contigs are represented as					
	* runs of N, but the exact sizes of the gaps are unknown.					
	* this record will be updated with the finished sequence					
	* as soon as it is available and the accession number will					
	* be preserved.					
	* 1 12248: contig of 12348 bp in length					
	* 12349 12448: gap of 100 bp					
	* 12449 13590: contig of 1142 bp in length					
	* 13591 13690: gap of 100 bp					
	* 13691 14194: contig of 1104 bp in length					
	* 14795 14894: gap of 100 bp					
	* 14895 16129: contig of 1235 bp in length					
	* 16130 16229: gap of 100 bp					
	* 16230 17330: contig of 1101 bp in length					
	* 17331 17430: gap of 100 bp					
	* 17431 18833: contig of 1033 bp in length					
	* 18834 18933: gap of 100 bp					
	* 18934 20519: contig of 1865 bp in length					
	* 20520 20619: gap of 100 bp					
	* 20620 20619: contig of 5578 bp in length					
	* 20619 20797: gap of 100 bp					
	* 20797 20812: contig of 9555 bp in length					
	* 20812 35992: gap of 100 bp					
	* 35993 46252: contig of 10233 bp in length					
BASE COUNT	57539	a	32376	c	32914	g
ORIGIN						
QUERY MATCH						
BEST LOCAL SIMILARITY	2.0%					
MATCHES	48.4%					
CONSERVATIVE	48.4%					
MISMATCHES	0					
INDELS	0					
GAPS	0					
QY	1891	agacatagttgttgcgttcacttgcgttgttca	1950			
DB	49689	AGACTTACTACTGATGTTATTCCTCTCGTATTGATGCTTATTCCTC	49748			
QY	1951	tttcggtttgttcggatcgatccatcgacttgctgtat	2010			





Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
1. (bases 1 to 153072)  
AUTHORS Wall, M.  
TITLE Direct Submission  
JOURNAL Submitted (22-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequests@sanger.ac.uk

COMMENT On Jan 26, 2001 this sequence version replaced 91-12227402. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TRIMBLE; Wp: WORMEP; Information on the WORMEP database can be found at:  
<http://www.sanger.ac.uk/Projects/Celebens/wormep> This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at:  
<http://www.sanger.ac.uk/HGP/Ch6>

RPL1-344F20 is from the library RPLR-11.2 constructed by the group of Pieter de Jong. For further details see:  
<http://www.chori.org/bacpac/home.htm>

VECTOR: PACB-6

IMPORTANT: This sequence is not the entire insert of clone RPL1-344F20. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap at 153072 in this sequence. The true right end of clone RPL1-344F20 is at 153053. The true right end of clone RPL1-505K1 is at 100 in this sequence.

FEATURES

source	Location/Qualifiers
	1. .153072
	/organism="Homo sapiens"
	/db_xref="Taxon:9606"
	/chromosome="6"
	/clone_id="RP11-344F20"
	/clone_start="RPCI-11.2"
repeat_region	/note="MIR repeat: matches 28. .144 of consensus"
repeat_region	/note="MIR repeat: matches 80. .252 of consensus"
repeat_region	/note="MIR repeat: matches 2641. .2717 of consensus"
repeat_region	/note="MIR repeat: matches 1. .239 of consensus"
repeat_region	/note="37 copies 2 mer ac 91% conserved"
repeat_region	/note="4732 repeat: matches 1. .310 of consensus"
repeat_region	/note="5207 repeat: matches 2 mer ta 69% conserved"
repeat_region	/note="6176 repeat: matches 33. .541 of consensus"
repeat_region	/note="63 copies 2 mer aa 59% conserved"
repeat_region	/note="6629. .7114 repeat: matches 2238. .2746 of consensus"
repeat_region	/note="7120. .7215 repeat: matches 2 mer ta 69% conserved"
repeat_region	/note="8404 repeat: matches 12. .365 of consensus"
repeat_region	/note="9833 repeat: matches 5184. .6143 of consensus"
repeat_region	/note="1184. .1211 repeat: matches 3692. .6120 of consensus"
repeat_region	/note="15150 repeat: matches 31. .250 of consensus"

repeat\_region /note="MIR repeat: matches 73. .218 of consensus" 15163. .15543

repeat\_region /note="MSTA repeat: matches 1. .426 of consensus" 15800. .15900

repeat\_region /note="MER33 repeat: matches 33. .135 of consensus" 15903. .16040

repeat\_region /note="L2 repeat: matches 2183. .2324 of consensus" 16258. .16489

repeat\_region /note="L2 repeat: matches 1818. .2033 of consensus" 16560. .16623

repeat\_region /note="M896 repeat: matches 1. .60 of consensus" 16918. .17030

repeat\_region /note="MER9IC repeat: matches 1. .119 of consensus" 17234. .17285

repeat\_region /note="26 copies 2 mer ag 88% conserved" 17305. .17526

repeat\_region /note="MIR repeat: matches 22. .258 of consensus" 18236. .18606

repeat\_region /note="LMD2 repeat: matches 594. .5968 of consensus" 18607. .18978

repeat\_region /note="LMD2 repeat: matches 551. .6321 of consensus" 20104. .20471

repeat\_region /note="LPA12 repeat: matches 4063. .4405 of consensus" 20472. .20828

repeat\_region /note="LPA12 repeat: matches 594. .5968 of consensus" 24508. .24802

repeat\_region /note="LPA12 repeat: matches 3. .371 of consensus" 20830. .22459

repeat\_region /note="MPEC-internal repeat: matches 4. .1578 of consensus" 20830. .22459

repeat\_region /note="MPEC repeat: matches 2287. .2287" 22463. .2287

repeat\_region /note="MPEC repeat: matches 1. .371 of consensus" 22828. .24196

repeat\_region /note="LPA12 repeat: matches 4405. .6153 of consensus" 24508. .24802

repeat\_region /note="LPA12 repeat: matches 3. .371 of consensus" 20830. .22459

repeat\_region /note="LPA15 repeat: matches 1. .285 of consensus" 25079. .25177

repeat\_region /note="LPA15 repeat: matches 1. .99 of consensus" 25187. .25625

repeat\_region /note="LPA15 repeat: matches 2128. .2598 of consensus" 25614. .25970

repeat\_region /note="LPA15 repeat: matches 5128. .5472 of consensus" 25971. .26271

repeat\_region /note="LPA15 repeat: matches 1. .300 of consensus" 26212. .26917

repeat\_region /note="LPA15 repeat: matches 5472. .6153 of consensus" 27112. .27147

repeat\_region /note="18 copies 2 mer tt 88% conserved" 27433. .27482

repeat\_region /note="25 copies 2 mer tt 76% conserved" 27489. .28551

repeat\_region /note="LIMAB repeat: matches 5254. .6288 of consensus" 28518. .28559

repeat\_region /note="THEIB repeat: matches 3. .364 of consensus" 28860. .30410

repeat\_region /note="THEIB-INTERNAL repeat: matches 1. .1580 of consensus" 30416. .30731

repeat\_region /note="THEIB repeat: matches 15. .364 of consensus" 30733. .31170

repeat\_region /note="LIMAB repeat: matches 4823. .5253 of consensus" 31622. .31890

repeat\_region /note="LIMEC repeat: matches 1738. .1988 of consensus" 31891. .32097

repeat\_region /note="MER6B repeat: matches 2. .208 of consensus" 32058. .32360

repeat\_region /note="LIMEC repeat: matches 1988. .2256 of consensus" 33222. .33283

repeat\_region /note="MER96 repeat: matches 115. .172 of consensus" 3410. .34490

repeat\_region /note="MER7A repeat: matches 1. .346 of consensus" 3838. .38597

repeat\_region /note="MER repeat: matches 31. .250 of consensus" 3838. .38597

repeat\_region /note="MIR repeat: matches 49. .124 of consensus" 41227. .41312



```

21164 21263: gap of 100 bp
21264 22640: contig of 1991 bp in length
22655 22754: gap of 100 bp
22755 24210: contig of 1336 bp in length
24291 24390: gap of 100 bp
24391 24594: contig of 1204 bp in length
25595 25694: gap of 100 bp
25695 27397: contig of 1703 bp in length
27398 27497: gap of 100 bp
27498 29331: gap of 100 bp
29331 29431: contig of 1033 bp in length
29431 31357: contig of 127 bp in length
31358 31457: gap of 100 bp
31458 33360: contig of 1903 bp in length
33361 33460: gap of 100 bp
33461 36064: contig of 2604 bp in length
36065 36164: gap of 100 bp
36164 37919: contig of 1315 bp in length
37980 38079: gap of 100 bp
38080 38080: contig of 2384 bp in length
40464 40563: gap of 100 bp
40564 43574: contig of 3011 bp in length
43575 45259: contig of 1585 bp in length
45260 45359: gap of 100 bp
45360 47305: contig of 2026 bp in length
47386 47485: gap of 100 bp
47486 49879: contig of 2194 bp in length
49880 49799: gap of 100 bp
49780 51819: contig of 2040 bp in length
51820 51919: gap of 100 bp
51920 55270: contig of 3351 bp in length
55271 55370: gap of 100 bp
55371 58223: contig of 2053 bp in length
58224 58323: gap of 100 bp
58324 61669: contig of 3346 bp in length
61570 61669: gap of 100 bp
61670 65776: contig of 4107 bp in length
65777 65876: gap of 100 bp
65877 68164: contig of 2288 bp in length
68165 68265: gap of 100 bp
68265 72331: contig of 4066 bp in length
72331 72430: gap of 100 bp
72431 75722: contig of 1391 bp in length
75722 75822: gap of 100 bp
75822 79328: contig of 3306 bp in length
79328 79428: gap of 100 bp
79428 86170: contig of 6743 bp in length
86170 86270: gap of 100 bp
86271 92478: contig of 6208 bp in length
92478 92579: gap of 100 bp
92579 98667: contig of 6069 bp in length
98668 98767: gap of 100 bp
98768 105029: contig of 6262 bp in length
105030 105129: gap of 100 bp
105130 111321: contig of 6192 bp in length
111422 111421: gap of 100 bp
111422 111747: contig of 6056 bp in length
117478 117578: gap of 100 bp
117578 125755: contig of 8178 bp in length
125755 128855: gap of 100 bp
128855 133433: contig of 7488 bp in length
133341 133443: gap of 100 bp
133443 141679: contig of 8236 bp in length
141680 141779: gap of 100 bp
141780 153666: contig of 11887 bp in length
153667 153765: gap of 100 bp
153767 162054: contig of 6868 bp in length
162053 162134: gap of 100 bp
162135 162135: contig of 374 bp in length.

1. 162508 Location/Qualifiers

Organism="Homo sapiens"

```

REFERENCE	AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, M., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bedal, F., Boeglukhin, V., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Cheopel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deverellano, K., Dewar, K., Domino, M., Doyle, M., Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gaze, D., Galagen, J., Gartland, S., Grant, G., Haiges, J., Heford, I., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lebocky, J., Levine, R., Liue, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McLean, P., McGurk, A., McKenna, N., McPhee, S., R., McLeod, J., Menous, L., Morrow, J., Nayor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rotman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange Thomas, N., Theodore, J., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A., and Zhou, M.	2 (bases 1 to 174200)
COMMENT	106-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA All repeats were identified using RepeatMasker: <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a>
TITLE	JOURNAL
FEATURES	source
Center: Whitehead Institute/ MIT Center for Genome Research	* 143089 143189 157621 157720: gap of 100 bp Center code: WIBR
Web site: <a href="http://www-seq.wi.mit.edu">http://www-seq.wi.mit.edu</a>	* 143189 157720: contig of 16480 bp in length Contact: sequence_submissions@genome.wi.mit.edu
Center project name: l6043	* 143189 157720: location/Qualifiers Center clone name: 281_G_13
----- Summary Statistics	1. 174200 /orgainism="Homo sapiens" <db_xref="taxon:9606" <clone="6" <chromosome="6" <map="6"
Sequencing vector: M13; M77815; 100% of reads	1. 3194 <note="assembly_fragment" <clone="RP11-281G13" <clone.lib="RPCI-11 Human Male BAC" <note="assembly_fragment" misc_feature
Chemistry: Dye-terminator Big Dye; 100% of reads	1. 3194 <note="assembly_fragment" misc_feature
Assembly program: Phrap version 0.960731	1. 3194 <note="assembly_fragment" misc_feature
Consensus quality: 163082 bases at least Q40	1. 3194 <note="assembly_fragment" misc_feature
Consensus quality: 168187 bases at least Q30	1. 3194 <note="assembly_fragment" misc_feature
Consensus quality: 170797 bases at least Q20	1. 3194 <note="assembly_fragment" misc_feature
Insert size: 182000; agarose-fp	1. 3194 <note="assembly_fragment" misc_feature
Insert size: 172100; sum-of-contigs	1. 3194 <note="assembly_fragment" misc_feature
Quality coverage: 3.6 in Q20 bases; agarose-fp	1. 3194 <note="assembly_fragment" misc_feature
Quality coverage: 3.8 in Q20 bases; sum-of-contigs	1. 3194 <note="assembly_fragment" misc_feature
-----	1. 3194 <note="assembly_fragment" misc_feature
* NOTE: This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	1. 3194 <note="assembly_fragment" misc_feature
* 3195 3294: gap of 3194 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 3295 5794: contig of 2500 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 5795 5894: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 5895 9861: contig of 3967 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 9862 9951: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 9962 13463: contig of 3503 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 13465 13564: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 13565 15205: contig of 1641 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 15206 15305: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 15305 19906: contig of 4604 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 19910 20009: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 20010 23963: contig of 3954 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 23964 24053: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 24054 29018: contig of 4955 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 29019 29118: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 29119 34619: contig of 5501 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 34620 34719: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature
* 34720 42050: contig of 7341 bp in length	1. 3194 <note="assembly_fragment" misc_feature
* 42051 42160: gap of 100 bp	1. 3194 <note="assembly_fragment" misc_feature



Db 50961 ARGRGMCKYTSCKYYSMFHKSSHDSSGGYYSRSYBVWTCMKDMSMCHSVDSHMMRD 50902  
 CC Sequencing vector: M13; M77815; 100% of reads  
 CC Chemistry: Dye-terminator Big dye; 100% of reads  
 CC Assembly program: Phrap; version 0.960731  
 CC Consensus quality: 151809 bases at least Q40  
 CC Consensus quality: 153486 bases at least Q30  
 CC Consensus quality: 154041 bases at least Q20  
 CC Insert size: 157000; agarose-fp  
 CC Insert size: 154641; sum-of-contigs  
 CC Quality coverage: 5.5 in Q20 bases; sum-of-contigs  
 CC Quality coverage: 5.6 in Q20 bases; sum-of-contigs  
 CC \* NOTE: This is a 'working draft' sequence. It currently  
 CC \* consists of 10 contigs. The true order of the pieces  
 CC \* is not known and their order in this sequence record is  
 CC \* arbitrary. Gaps between the contigs are represented as  
 CC \* runs of N, but the exact sizes of the gaps are unknown.  
 CC \* This record will be updated with the finished sequence  
 CC \* as soon as it is available and the accession number will  
 CC \* be preserved.  
 CC \* 1 3918: contig of 3918 bp in length  
 CC \* 3919 4018: gap of 100 bp  
 CC \* 4019 7852: contig of 3834 bp in length  
 CC \* 7853 7952: gap of 100 bp  
 CC \* 7953 10420: contig of 2468 bp in length  
 CC \* 10421 10520: gap of 100 bp  
 CC \* 10521 16131: contig of 5611 bp in length  
 CC \* 16132 16231: gap of 100 bp  
 CC \* 16232 21713: contig of 5482 bp in length  
 CC \* 21714 21813: gap of 100 bp  
 CC \* 21814 25983: contig of 4170 bp in length  
 CC \* 25984 33603: gap of 100 bp  
 CC \* 33605 33704: gap of 100 bp  
 CC \* 33705 45042: contig of 1138 bp in length  
 CC \* 45043 45142: gap of 100 bp  
 CC \* 45143 77339: contig of 32197 bp in length  
 CC \* 77340 77439: gap of 100 bp  
 CC \* 77440 15542: contig of 78103 bp in length.  
 XX FH Key Location/Qualifiers  
 [2] FH source  
 RP source 1. .15542  
 RA Birren B., Linton L., Nusbaum C., Lander E., Allen N., Anderson M.,  
 RA Baldwin J., Barna N., Beckerly R., Boukhalal B.,  
 RA Brown A., Castle A., Colangelo M., Collins S., Collymore A., Cooke P.,  
 RA Dearleano K., Dewar K., Domino M., Donegan L., Boyle M., Ferreira P.,  
 RA Fitzhugh W., Forrest C., Funke D., Gade D., Galaaen J., Gardina S.,  
 RA Grant G., Hagos B., Heaford A., Horton L., Howland J.C., Johnson R.,  
 RA Jones C., Kann L., Karatas A., Klein J., Lebocky J., Lieu C., Locke K.,  
 RA Macdonald P., Marquis N., McEvane P., McGuirk A., McKernan K., McLaughlin J.,  
 RA Melior J., Morrow J., Nayor J., Norman C.H., O'Connor T., O'Donnell P.,  
 RA Peterson J., Pollicra V., Riley R., Roy A., Santos R., Severy P.,  
 RA Strange-Thomson N., Stojanovic N., Subramanian A., Talama J., Tesfave S.,  
 RA Tirrell A., Vassilieff H., Wo A., Wheeler J., Wu X., Wyman D., Ye W.J.,  
 RA Zimmer A., Zody M.;  
 RT ; Submitted (13-Nov-1999) to the EMBL/GenBank/DDBJ databases.  
 RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,  
 RL Cambridge, MA 02141, USA  
 XX All repeats were identified using RepeatMasker:  
 CC All repeats were identified using RepeatMasker:  
 CC Smil, A.P.A. & Green, P. (1996-1997)  
 CC http://ftp.genome.washington.edu/RW/RepeatMasker.html  
 CC Center: Whitehead Institute/ MIT Center for Genome Research  
 CC Center code: WIR  
 CC Web site: http://www-seq.wi.mit.edu  
 CC Contact: sequence.submissions@genome.wi.mit.edu  
 CC Center project name: I3101  
 CC Center clone name: 10\_L7  
 CC ----- Summary Statistics  
 SQ Sequence 15542 BP: 44570 A; 32359 C; 30817 G; 46895 T; 901 other;  
 CC Query Match 1 - 8%; Score 38.6; DB 30; Length 15542;  
 CC Best Local Similarity 51.4%; Pedi. No. 9.7; Mismatches 0; Gaps 0;  
 CC Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;





Search completed: September 15, 2002, 23:57:46  
Job time: 8735 sec

Copyright (c) 1993 - version 4.5 GenCore - nucleic search, using sw model compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 15, 2002, 22:18:56 : Search time 252.83 seconds

(without alignments)  
14783.541 Million cell updates/sec

Title: US-09-828-313-22  
Perfect score: 2177  
Sequence: 1 atcccgaggctgttattggct.....gaagatattccgcgttaacg 2177  
Scoring table: TDNNTRY.NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	N_GeneSeq_0328022:*
1:	/SIDS1/gcadata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1981.DAT:*
3:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1982.DAT:*
4:	/SIDS1/gcadata/hold-geneseqn-emb1/NA1983.DAT:*
5:	/SIDS1/gcadata/hold-geneseqn-emb1/NA1984.DAT:*
6:	/SIDS1/gcadata/hold-geneseqn-emb1/NA1985.DAT:*
7:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1986.DAT:*
8:	/SIDS1/gcadata/hold-geneseqn-emb1/NA1987.DAT:*
9:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1988.DAT:*
10:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1989.DAT:*
11:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1990.DAT:*
12:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1991.DAT:*
13:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1992.DAT:*
14:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1993.DAT:*
15:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1994.DAT:*
16:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1995.DAT:*
17:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1996.DAT:*
18:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1997.DAT:*
19:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1998.DAT:*
20:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA1999.DAT:*
21:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA2000.DAT:*
22:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA2001.DAT:*
23:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA2002.DAT:*
24:	/SIDS1/gcadata/hold-geneseqn-geneseqn-emb1/NA2003.DAT:*

Pre-d. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	2177	100.0	2177	2A	ABA91077 Physcomitrella pat
2	540.2	24.8	683	2A	ABA91064 Physcomitrella pat
c	35.4	5059	20	AAX84322	Steinbush virus nuc
c	34.8	5845	24	ABJ3301	Human immune syste
c	34.6	1.6	2625	2ABJ15017	Human nervous syst
c	34.6	1.6	3612	2A251221	N. tabacum Nhl1 ge
7	34.4	1.6	7201	2ABL05860	Drosophila melanog
c	34.4	1.6	7581	2ABL05874	Drosophila melanog
c	9	1.6	14458	23 AB10220	Drosophila melanog

#### SUMMARIES

**RESULT** 1  
 ABA1077 ID- ABA91077 standard; cDNA; 2177 BP.  
 XX AC ABA91077;  
 XX DT 22-FEB-2002 (first entry)  
 XX DE Physcomitrella patens MPK-3 full-length cDNA, SEQ ID NO:22.  
 XX KW Protein kinase stress-related protein; PKRP; moss; protein kinase; PK-6; protein kinase-7; PK-7; protein kinase-8; protein kinase-9; casein kinase homologue-1; CK-1; cassin kinase homologue-2; casein kinase homologue-3; CK-3; mitogen-activated protein kinase; MAP kinase-2; MPK-2; MAP kinase-3; MPK-3; MAP kinase-4; MPK-4; MAP kinase-5; MPK-5; calcium-dependent protein kinase-1; CPK-1; CPK-2; oxorepression; calcium-dependent protein kinase-2; CPK-2; oxorepression; environmental stress; saltinity; drought; temperature; tolerance; transgenic plant; EST; expressed sequence tag; ss.  
 XX OS Physcomitrella patens.  
 XX PN WO200177356-A2.  
 XX PD 18-OCT-2001.  
 XX PR 06-APR-2001; 2001WO-US11435.  
 XX PR 07-APR-2000; 2000US-196001P.  
 PA (BADI ) BASF PLANT SCI GMBH.  
 XX



QY	1501	ctccaaacttttgtcttcataaaggatctccaaatggccgtatgttca	1560	xx
DB	1501	cctccacattttgtcttcataaaggatctccaaatggccgtatgttca	1560	xx
PR	06-APR-2001; 2001WO-US1435.			
XX				
PR	07-APR-2000; 2000US-19601P.			
XX				
PA	(BADI ) BASE PLANT SCI GMBH.			
XX				
PI	Da Costa Silva EO, Bohnert HJ, Van Thiel N, Chen R;			
PT	Sarría-Millan R;			
XX				
DR	WPI: 2002-049153/06.			
XX				
PT	New protein, useful for increasing tolerance to environmental stress, comprises a Protein Kinase Stress-Related Protein selected from protein kinases, Casein kinase homologs, MAP kinases or Calcium dependent protein kinases -			
XX				
PS	Example 5, Fig 11; 154pp; English.			
XX				
CC	Sequences AAM52810-AAM52842 represent novel protein kinase stress-related proteins (PKSRPs) from the moss Physcomitrella Patens, and sequences CC			
CC	ABA91069 ABA91068 represent full-length cDNAs encoding them. The cDNA CC			
CC	sequences were obtained from expressed sequence tags (ESTs; ABA91056-ABA91068) derived from Physcomitrella patens cDNA libraries. The PKSRPs CC			
CC	of the invention comprise protein kinase-6 (PK-6), protein kinase-7 (PK-7), protein kinase-8 (PK-8), protein kinase-9 (PK-9), casein kinase CC			
CC	homologue-1 (CK-1), casein kinase homologue-2 (CK-2), casein kinase CC			
CC	homologue-3 (CK-3), mitogen-activated protein (MAP) kinase-2 (MKR-2), MAP kinase-3 (MPK-3), MAP kinase-4 (MPK-4), MAP kinase-5 (MKR-5), CC			
CC	calcium-dependent protein kinase-1 (CPK-1), and calcium-dependent protein CC			
CC	kinase-2 (CPK-2). When overexpressed, the PKSRPs are able to confer CC			
CC	tolerance to environmental stresses such as salinity, drought, temperature, metal, chemical, pathogenic and oxidative stress.			
CC	Physcomitrella Patens PKRP nucleic acids may be used to generate transgenic plants and seeds with increased tolerance to salinity, drought CC			
CC	and temperature. The transgenic plants generated can be monocots or CC			
CC	dicots and are especially maize, wheat, rye, oat, triticale, rice, barley, cotton, rapeseed, cassava, sunflower, tagetes, leguminous plants CC			
CC	(e.g., soybean, peanut, Vicia species, alfalfa), solanaceous plants CC			
CC	(e.g., potato, tobacco,aubergine, pepper, tomato), coffee, cacao, tea, CC			
CC	Salix species, oil palm, coconut, perennial grasses and forage crops. The CC			
CC	PKSRP nucleotide and proteins may also be used in evolutionary and CC			
CC	protein structural studies and as markers for specific regions of the genome.			
XX	.			
SQ	Sequence 683 BP; 172 A; 188 C; 163 G; 158 T; 2 other;			
Query Match	24.8%	Score 540.2; DB 24; Length 683;		
Best Local Similarity	96.6%	Pred. No. 1e-172;		
Matches	627;	Conservative		
		Mismatches 0;		
		Indels 13;		
		Gaps 9;		
		7;		
QY	868	ccctcgatggagacgacatcgaaacccatggatcgatgtttcatcgccatgtcg 927		
DB	10	cctcgatggagacgacatcgaaacccatggatcgatgtttcatcgccatgtcg 69		
QY	928	agcacttcgatccatcgaaatcgatccatcgaaatcgatccatcgaaatcgatccatcg 987		
DB	70	agcacttcgatccatcgaaatcgatccatcgaaatcgatccatcgaaatcgatccatcg 129		
QY	988	cggaggccatcgccgcggactccaaaggatattcgatcgatccatcgatccatcg 1047		
DB	130	cggaggccatcgccgcggactccaaaggatattcgatcgatccatcgatccatcg 189		
QY	1048	aatcgccgcggactccaaatcgatccatcgatccatcgatccatcgatccatcg 1107		
DB	190	aatcgccgcggactccaaatcgatccatcgatccatcgatccatcgatccatcg 249		
QY	1108	aatcgccgcggactccaaatcgatccatcgatccatcgatccatcgatccatcg 309		
DB	250	aatcgccgcggactccaaatcgatccatcgatccatcgatccatcgatccatcg 309		

```

QY 1168 gaaggctgtcatggatccactccatcgatgttccatccatcgatgtttcccccggaa 1227
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| FT
Db 310 gaacgcgtcgtcaatggatccatccatcgatgtttcccccggaa 369 FT
FT misc_difference 3642 /*tag*/ h "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
Db 370 gaacgcgtcgtcaatggatccatccatcgatgtttcccccggaa 429 FT
FT misc_difference 3657 /*tag*/ i "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
QY 1288 tgcgtcccaaggacgacgttcggccaaacatccacccatcgatgtttcccccggaa 1345
Db 430 tgcgtcccaaggacgacgttcggccaaacatccacccatcgatgtttcccccggaa 489 FT
FT misc_difference 3659 /*tag*/ j "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
QY 1346 ttgttagtcggactcgccatcgatgtttcccccggaa 1404 FT
Db 490 ttgttagtcggactcgccatcgatgtttcccccggaa 609 FT
FT misc_difference 3681 /*tag*/ k "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
QY 1462 attcccccatacgatgtttcccccggaa 1510 FT
Db 610 at--cccatcgatgtttcccccggaa 655 FT
FT misc_difference 3698 /*tag*/ l "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3703 /*tag*/ m "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
DE 08-SEP-1999 (first entry) FT
XX Steealth virus nucleic acid clone, SEQ ID: 24. FT
XX Stealth virus; detection; diagnosis; infection; ss. FT
XX OS Stealth virus. FT
PH misc_difference 3605 /*tag*/ o "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3713 /*tag*/ p "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3720 /*tag*/ q "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3723 /*tag*/ r "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3732 /*tag*/ s "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3740 /*tag*/ t "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3747 /*tag*/ u "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3750 /*tag*/ v "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"
FT misc_difference 3751 /*tag*/ w "this nucleotide is represented as a * in the
FT /note* specification, and is included to maintain the base numbering given in the specification"

```

```

FT      /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3766 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3778 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3799 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3800 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3812 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3813 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3818 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3819 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3824 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3827 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3841 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3853 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3857 /*tag= */ "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"

```



PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0228287.  
 PR 01-SEP-2000; 2000US-0228343.  
 PR 01-SEP-2000; 2000US-0228344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232801.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 23-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0235488.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 11-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 08-NOV-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 08-NOV-2000; 2000US-0241826.  
 PR 20-OCT-2000; 2000US-0242221.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0246207.  
 PR 17-NOV-2000; 2000US-0246208.  
 PR 17-NOV-2000; 2000US-0246209.  
 PR 17-NOV-2000; 2000US-0246210.  
 PR 17-NOV-2000; 2000US-0246211.  
 PR 17-NOV-2000; 2000US-0246212.  
 PR 17-NOV-2000; 2000US-0246213.  
 PR 17-NOV-2000; 2000US-0246214.  
 PR 17-NOV-2000; 2000US-0246215.  
 PR 17-NOV-2000; 2000US-0246216.  
 PR 17-NOV-2000; 2000US-0246217.  
 PR 17-NOV-2000; 2000US-0246218.  
 PR 17-NOV-2000; 2000US-0246244.  
 PR 17-NOV-2000; 2000US-0246264.  
 PR 17-NOV-2000; 2000US-0246265.  
 PR 17-NOV-2000; 2000US-0246266.  
 PR 17-NOV-2000; 2000US-0246267.  
 PR 17-NOV-2000; 2000US-0246268.  
 PR 01-DEC-2000; 2000US-0251160.  
 PR 01-DEC-2000; 2000US-0251160.  
 PR 03-DEC-2000; 2000US-0251030.  
 PR 03-DEC-2000; 2000US-0251088.  
 PR 05-DEC-2000; 2000US-0251679.  
 PR 05-DEC-2000; 2000US-0251779.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251888.  
 PR 08-DEC-2000; 2000US-0251889.  
 PR 08-DEC-2000; 2000US-0251980.  
 PR 08-DEC-2000; 2000US-0254097.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-541565/60.  
 XX DR  
 XX Disclosure; SEQ ID NO 7348; 1701pp + Sequence Listing; English.  
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -  
 PT  
 XX Disclosure; SEQ ID NO 7348; 1701pp + Sequence Listing; English.  
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -  
 CC The invention relates to novel genes (ABA11004-ABA21534) and Proteins (ABA14678-ABA18011) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (anti)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or uterine; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
 CC Sequence 2625 BP; 699 A; 581 G; 701 T; 0 other;  
 XX











CC  
CC  
XX  
SQ  
Sequence 8375 BP; 2190 A; 134 C; 2005 G; 4046 T; 0 other:  
rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

Search completed: September 15, 2002, 23:46:06  
Job time: 5230 sec

1



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

Om nucleic - nucleic search, using sw model

Run on : September 15, 2002, 21:47:06 ; Search time 59.1 Seconds  
(without alignments)  
9048.124 Million cell updates/sec

Title: US-09-828-313-22  
perfect score: 2177  
sequence: 1 atcccggttgcgtatgt.....gcagatattgcgttaacgc 2177

Scoring table: IDENTITY\_NUC  
GapPenalty 0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : [ISSUED\\_PATENTS\\_NA\\_\\*](#)

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/seq: \*  
4: /cgn2\_5/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/nackilles1.seq: \*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	score	Query Match Length	DB ID	Description
1	41.6	1.9	7218 1	US-08-232463-14
2	31.8	1.5	1935 2	US-08-878-989-11
3	31.8	1.5	1935 4	US-09-272-796-11
4	31.8	1.5	4403765 4	US-09-103-8404-2
5	31.8	1.5	44110529 4	US-09-103-8404-1
6	31.2	1.4	396 3	US-08-963-168C-4
7	31.2	1.4	458 3	US-09-141-000-4
8	30.4	1.4	1101 4	US-09-078-294-20
9	30.4	1.4	1791 1	US-08-445-94-7
10	30.4	1.4	1791 1	US-08-474-499-7
11	30.4	1.4	1791 1	US-08-307-279A-7
12	30.4	1.4	1791 5	PCM-S95-06221-7
13	30.4	1.4	4164 1	US-08-204-675-1
14	30.4	1.4	4164 2	US-08-660-754-1
15	30.4	1.4	4164 2	US-08-796-364-1
16	30.4	1.4	4164 5	PCM-US95-02520-1
17	30.4	1.4	4933 1	US-08-204-675-2
18	30.4	1.4	4933 2	US-08-660-754-2
19	30.4	1.4	4933 2	US-08-796-364-2
20	30.4	1.4	4933 5	PCM-US95-02520-2
21	30.4	1.4	5181 3	US-08-801-444-5
22	30.4	1.4	5181 4	US-09-498-599-5
23	30.4	1.4	7102 3	US-09-138-024-20
24	30.4	1.4	7333 3	US-09-138-024-21
25	30.4	1.4	7633 3	US-09-028-831-1
26	30.4	1.4	7633 4	US-08-815-500-1
27	30.4	1.4	7633 4	US-09-273-163-1

RESULT 1  
US-08-232463-14  
; Sequence 14, Application US/08232463.  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHIFFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT POMPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Hardier  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BERNI, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)935-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZ9PT-F1S  
US-08-232463-14

```

; LENGTH: 1935 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORATO4
; CLONE: 705365
; US-08-878-981-11

Query Match          1.5%; Score 31.8; DB 22
Best Local Similarity 51.8%; Pred. No. 2.7;
Matches 72; Conservative 51.8%; 0; Mismatches 62
GENERAL INFORMATION:
APPLICANT: Bandman, Olga L.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Gugeler, Karl G.
APPLICANT: Lal, Preeti
APPLICANT: Goil, Surya K.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: DISEASE ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 31740 Porter Drive
CITY: Palo Alto
STATE: CA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Disette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/272,796
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/878,989
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billing, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0321 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-825-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1935 base pairs
TYPE: nucleic acid

```



```

; LENGTH: 396 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE: 
; NAME/KEY: Coding Sequence
; LOCATION: 1...393
; US-08-953-168C-4

Query Match 1.4%; Score 31.2; DB 3; Length 396;
Best Local Similarity 45.8%; Pred. No. 1.5; Matches 108; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
Db 339 GCGGCTTCCATTCGGGAGAACCGCATTTGCCCTCCATTCGGAGAAAGGGCTCTCAT 1179
Qy 1120 gcggggttaagtgcggtggaaacggcgttttgtcgagacatccggatggatggggcttcgtcat 1179
Db 1180 ggaatcccaactccatgttttccggataggtagtggtctcgccggaaagggcttcac 1239
Qy 279 ACCATGCCATTCGGATCCACCTTCGCCGATCCGGAGAACCTCCATTC 220
Db 1240 cccatccctatctggccggaaacatttcacccggccacggaaatgtgggtgtcccaaaag 1299
Qy 219 ACCGAATTCCGCCTGACCTCACCCACTCCGAAGGCCATATCCCGGATCC 160
Db 1300 gagcgttccctcccccggaaacgttccacccatccccacccatgttgccgc 1355
Qy 159 GCCTTACCTCCGCCATTCGGCATACGCCAATCCSCTGCGTCATCCGC 104
Db ; SH.CN.DENKASSRSCSTHSSCC.SDDNEHS 51

RESULT 7
US-09-41-000-4/c
; Sequence 4, Application US/09141000
; PATENT NO.: 6054295
; GENERAL INFORMATION:
; APPLICANT: Chen, Fang
; TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NUCLEAR
; FILE REFERENCE: 19999Y
; CURRENT APPLICATION NUMBER: US/09/141,000
; CURRENT FILING DATE: 1998-08-26
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 4
; LENGTH: 458
; TYPE: PR
; ORGANISM: Human
; US-09-141-000-4

Query Match 1.4%; Score 31.2; DB 3; Length 458;
Best Local Similarity 45.8%; Pred. No. 1.6; Matches 34; Conservative 121; Mismatches 237; Indels 0; Gaps 0;
Db 412 tatcttgactgtttccatggggccacgttaacccatggatccgc 471
Qy 472 gtacatttttacttcgaaagacaaaaatttgcacatgt 515
Db ; US-09-141-000-4

RESULT 8
US-09-78-294-20
; Sequence 20, Application US/09078294
; PATENT NO.: 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 20
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: BAC-F2 contig 47
; US-09-078-294-20

Query Match 1.4%; Score 30.4; DB 4; Length 1101;
Best Local Similarity 55.8%; Pred. No. 5.6; Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
Db 2124 gttttatataactatgtggaaacggctgtggatgtt 2167
Qy 472 gtacatttttacttcgaaagacaaaaatttgcacatgt 515
Db ; US-09-78-294-20

RESULT 9
US-08-245-294-7/c
; Sequence 7, Application US/08245294
; PATENT NO.: 5644047
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regenry, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOsing ROCHALIMEA HENSELAE AND ROCHALIMEA QUINTANA
; TITLE OF INVENTION: INFECTIOn
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOSS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.25
; APPLICATION APPLICATION DATA:
; FILING DATE: US/08/245,294

```

CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Spratt, Gwendolyn D.  
 REFERENCE/DOCKET NUMBER: 36.016  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 404/688-0770  
 TELEFAX: 404/688-9880  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1791 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 141..1649  
 ; US-08-245-294-7  
 ;  
 RESULT 10  
 US 08 474 499-7/C  
 Sequence 7, Application US/08474499  
 ; Patent No. 569376  
 GENERAL INFORMATION:  
 APPLICANT: Anderson, Burt E.  
 APPLICANT: Regney, Russell L.  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
 DIAGNOSING ROCHALIMAEA HENSELAE  
 TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
 STREET: 127 Peachtree Street, Suite 1200  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30303  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/474,499  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/245, 294  
 FILING DATE: 18-MAY-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Spratt, Gwendolyn D.  
 REGISTRATION NUMBER: 36.016  
 REFERENCE/DOCKET NUMBER: 1414.624  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404) 688-0770  
 TELEFAX: (404) 688-9880  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1791 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 141..1649  
 ; US-08-474-499-7  
 ;  
 RESULT 11  
 US-08-307-279A-7/C  
 Sequence 7, Application US/08307279A  
 ; Patent No. 578347  
 GENERAL INFORMATION:  
 APPLICANT: Anderson, Burt E.  
 APPLICANT: Regney, Russell L.  
 TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae  
 TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
 STREET: 127 Peachtree Street, N.E., Suite 1200  
 CITY: Atlanta  
 STATE: Georgia  
 COUNTRY: USA  
 ZIP: 30303  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/307, 279A  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Spratt, Gwendolyn D.  
 REGISTRATION NUMBER: 36.016  
 REFERENCE/DOCKET NUMBER: 1414.624  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404) 688-0770  
 TELEFAX: (404) 688-9880  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1791 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)





Mon Sep 16 08:49:19 2002

us-09-828-313-22.rni

Page 8

Copyright GenCore version 4.5  
(c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 15, 2002, 21:28:31 ; search time 1744.24 seconds  
(w/without alignments)  
16345.658 Million cell updates/sec

**title:** US-09-828-313-22  
**perfect score:** 2177  
**sequence:** 1 atcccggtgtatggct.....gcaagatattgcgttaacgc 2177

**Scoring table:** IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

**Searched:** 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

**Database :**

```

EST.*:
1: em_estba:*
2: em_estbun:*
3: em_estbin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
1	290.8	13.4	294	10	BJI79058	BJI79058	RESULT 1
2	202.8	9.3	547	10	BJI74514	BJI74514	LOCUS BJI79058
3	113.6	5.2	605	10	BJI74924	BJI74924	DEFINITION 294 bp mRNA linear EST 24-JAN-2002
4	50.4	2.3	525	9	AW739415	gba1f06.y	caulonemata and malformed buds Physcomitrella patens
5	42	1.9	855	12	CNS011PE	ALI00556	cDNA clone pphb2aa07 5', mRNA sequence.
6	39.8	1.8	949	12	AG029463	AG029463	VERSION BJI79058.1 G1:18347012
7	39.2	1.8	1101	12	CNS00100	BJI79058	KEYWORDS EST.
8	38.6	1.8	553	9	AM625386	AL068607	Physcomitrella patens subsp. patens.
9	38.6	1.8	671	9	AW930530	AW625386	Eukaryota; Viriplantae; Streptophyta; Embryophyta; Bryophyta;
10	37.8	1.7	577	12	AG155394	AW930530	Fungiidae; Funariales; Funariaceae; Physcomitrella.
11	37.4	1.7	513	10	BG367216	AW930530	REFERENCE 1 (bases 1 to 294)
12	37.4	1.7	940	10	BP023571	AL073444	AUTHORS Fujita, T., Shin-i, T., Seii, M., Kamiya, A., Uchiyama, I., Nishiyama, T., M.
13	37.4	1.7	1001	12	CNS00167	BP023571	TITLE Comparison of the moss Physcomitrella patens genome with flowering
14	37.2	1.7	513	12	B87304	AL073444	CONTACT: Tadashi Shin-i
15	37.2	1.7	1067	12	CNS00269	BP023571	JOURNAL Unpublished (2002)
16	37	1.7	507	9	A1775253	AL073444	COMMENT Center For Genetic Resource Information
							National Institute of Genetics
							111 Yata, Mishima, Shizuoka 411-8540, Japan
							Tel: 81-559-816556
							Fax: 81-559-81-6655
							Email: tshini@genes.nig.ac.jp
							Background of the vector is pBluescript II, that was <i>in vivo</i>
							excised from a modified λ phage vector (Mo Bi Tec, Germany). XhoI
							digested-5' end of cDNA is ligated to Sall site of the vector, and
							the BamHI digested-5' end including poly-A tail is ligated to BamHI
							site of the vector. cDNA insert could be amplified with
							conventional T7 and T3 primers. This normalized full-length cDNA
							library was generated basically according to the method described
							in Genome Research 10, 1617-1630 (2000), Carninci, P. et al.
							Prokaryotes were blended with the polyTNT, and then cultivated on
							the BD medium containing 0.5 μM BA (benzylaminopurine) for 8 to 13
							days under the continuous light.
							Location/Qualifiers
							1. .294
							FEATURES Source















REFERENCE  
2 (bases 1 to 1067)  
AUTHORS Roest-crollius,H., Jaillon,O., Dassila,C., Bouneau,L., Fisher,C.,  
Benoit,A., Flizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.  
TITLE Human gene number estimate provided by genome wide analysis using  
JOURNAL Tetraodon nigroviridis DNA sequence  
REFERENCE Unpublished  
AUTHORS 3 (bases 1 to 1067)  
TITLE Genoscope.  
JOURNAL Direct Submission  
COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.  
FEATURES Location/Qualifiers  
source  
1. ..1067  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:9983"  
/clone="021D17"  
/clone\_id="G"  
/note="Genoscope sequence ID : c0BG021CB09LP1-end : T7"  
BASE COUNT 356 a 185 c 268 g 257 t 1 others  
ORIGIN

Query Match 1 7%; Score 37.2; DB 12; Length 1067;  
Best Local Similarity 54.8%; Pred. No. 14;  
Matches 75; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
Matches 75; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
Oy 1702 tggtatcccttggcgacatctttgttcggcaggttttcaagttttggtttgt 1761  
Db 624 TGATTTATGTGACTCAGTCAGTGCTGGCGTCGCCCGCTTTAAATTGATGTTGTTGCT 565  
Oy 1762 aggggccttagggactatacgaaatggaaatgtgtgggtggatttgaccaa 1821  
Db 564 GGGGTCAGTGAATGTTGTCAGGAATGATACATGTTGGTGCATATATAGAA 505  
Oy 1822 tcrtttggggtttgt 1839  
Db 504 TAGTTGTTGAAAGTT 487

Search completed: September 15, 2002, 22:53:08  
Job time: 5077 sec





XX Disclosure; SEQ ID NO 36795; 21pp + Sequence Listing; English.

PS XX

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB13051), expressed DNA sequences (ABU01840-ABU16175) and the encoded proteins (ABR57737-ABY2072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pdb/published\\_pct Sequences](http://wipo.int/pdb/published_pct Sequences).

CC XX

SQ Sequence 1381 AA;

Query Match Similarity 7.2%; Score 125.5; DB 22; Length 1381; Matches 72; Conservative 46; Mismatches 135; Index 69; Gaps 15; Oy 30 SCSVPHDLGLGNP-DYYQGSTPYITNRMAAEVL----EYGVTYFCATPNQPFLLER 83 Db 367 stthihalidegsekrykaenpygsspsgeoseflidpkgrktksarrlagdekk 426 Qy 84 KQKVWVKGSKPRSKKGKSVNRELSRSPHGSQSQSKVDFYHSTMQOLESRINSGPEH 143 Db 427 -ptkgggmvprr-----rktspprpadast-----gnspkqspakr-pkpq 471 Qy 144 LAGDRDSKHKLKVFRTRLAURALPRISIDNEESPLNPNSGEALSPRATAKDETQML 202 Db 472 ldnneqlhemysytsdhali-----mdtrkspslnadsknptdsnsplnlti 523 Qy 203 KSARSBLGVMVSKRQEYFLRRRKFAWKVLUOSISEMKPKWMEHTPMAYRSRGSP- 261 Db 524 vtttispldkspv-----spkaprtptaspkifnp---kpsgspk 561 Qy 262 -----PKRASPSLPGPKNIS--PPRQVSVPOTSSPKKNNSP-PPQPAFVA----R 306 Db 562 drspvqpkppaplpqrppmetpkakspnprksispnnkrqppplrksprrpepkv-pallr 621 Qy 307 TASKYS-AASQVQVRNNGAKS 327 Db 622 saepvqlgvnqnvrfsqqtis 643

RESULT 2

AAR85442 ID AAR85442 standard; Protein; 912 AA.

XX AC AAR85442;

XX DT 19-FEB-1996 (first entry)

DE Bovine brevican core protein.

XX KW Brevican; chondroitin sulphate proteoglycan; glial cell; axon; neurofibromatosis; gliosis.

XX OS *Bos taurus*.

Key Location/Qualifiers

FT 1..34 Peptide /label= sig\_peptide

FT Region /label= Immunoglobulin-like\_loops

FT 35..158 Modified-site /label= N-glycosylation\_site

FT 130 Region /label= Protein-like\_tandem\_repeats

FT 337 Modified-site /label= N-glycosylation\_site

Query Match Similarity 7.2%; Score 125.5; DB 22; Length 1381; Matches 72; Conservative 46; Mismatches 135; Index 69; Gaps 15; Oy 30 SCSVPHDLGLGNP-DYYQGSTPYITNRMAAEVL----EYGVTYFCATPNQPFLLER 83 Db 367 stthihalidegsekrykaenpygsspsgeoseflidpkgrktksarrlagdekk 426 Qy 84 KQKVWVKGSKPRSKKGKSVNRELSRSPHGSQSQSKVDFYHSTMQOLESRINSGPEH 143 Db 427 -ptkgggmvprr-----rktspprpadast-----gnspkqspakr-pkpq 471 Qy 144 LAGDRDSKHKLKVFRTRLAURALPRISIDNEESPLNPNSGEALSPRATAKDETQML 202 Db 472 ldnneqlhemysytsdhali-----mdtrkspslnadsknptdsnsplnlti 523 Qy 203 KSARSBLGVMVSKRQEYFLRRRKFAWKVLUOSISEMKPKWMEHTPMAYRSRGSP- 261 Db 524 vtttispldkspv-----spkaprtptaspkifnp---kpsgspk 561 Qy 262 -----PKRASPSLPGPKNIS--PPRQVSVPOTSSPKKNNSP-PPQPAFVA----R 306 Db 562 drspvqpkppaplpqrppmetpkakspnprksispnnkrqppplrksprrpepkv-pallr 621 Qy 307 TASKYS-AASQVQVRNNGAKS 327 Db 622 saepvqlgvnqnvrfsqqtis 643

RESULT 2

AAR85442 ID AAR85442 standard; Protein; 912 AA.

XX AC AAR85442;

XX DT 19-FEB-1996 (first entry)

DE Bovine brevican core protein.

XX KW Brevican; chondroitin sulphate proteoglycan; glial cell; axon; neurofibromatosis; gliosis.

XX OS *Bos taurus*.

Key Location/Qualifiers

FT 1..34 Peptide /label= sig\_peptide

FT Region /label= Immunoglobulin-like\_loops

FT 35..158 Modified-site /label= N-glycosylation\_site

FT 130 Region /label= Protein-like\_tandem\_repeats

FT 337 Modified-site /label= N-glycosylation\_site

Query Match Similarity 6.7%; Score 116; DB 16; Length 912; Matches 54; Conservative 24; Mismatches 101; Index 101; Gaps 42; Oy 140 PEPFLAGDRDSKHKLKVFRTRLAURALPR-----ISIDME----SPLPNIS 183 Db 368 pashlasdate--aivvtetleekltpqavevesergalysipledgggstspped 424 Qy 184 GEALSPRATAKDETQMLKAASSELGNYVKROFYLRARRRK-----FA 232 Db 425 aea- -pttliefetqslvpp1gsseeqkvieqky-geekkeeeeeeevedeala 481 Qy 233 WKEVYLOSISENEMKPKWMEHTPMAYRSRGSPKNA----STPSLPGKNTSPPROVSFQ 286 Db 482 wpselssldpeaplpptpyppeesitqaapppvralqgyrspppydeapepprprviggpt 541 Qy 287 RSPPPKK-----VSPPPQPAFARTISKSAASQVQVRNNGAK 326 Db 542 kelpptpregnlasappstlvgarereieetggpelsgaprgese 584

RESULT 3

ID AAB20062 ID AAB20062 standard; Protein; 2001 AA.

XX AC AAB20062;

XX DT 23-APR-2001 (first entry)

XX DE Arabidopsis thaliana silencing gene-encoded protein.

XX KW Gene silencing; silencing gene; MOM.

XX OS *Arabidopsis thaliana*.

XX Key Location/Qualifiers

FT	Region	177..350	
FT	Peptide	/note= "internal repeat region"	
FT	Binding-site	/note= "nuclear localization motif"	
FT	Region	479..719	
FT	Misc-difference	705	
FT	Peptide	/note= "Lys in ecotype Columbia"	
FT	Peptide	832..838	
FT	Domain	838..862	
FT	Misc-difference	1219	
FT	Region	1462..1672	
FT	Region	1948..1894	
FT	Region	/note= "internal repeat region"	
FT	Domain	1899..1941	
FT	Misc-difference	1219	
FT	Region	1462..1672	
FT	Region	1948..1894	
FT	Region	/note= "internal repeat region"	
FT	Domain	1899..1941	
FT	PR	21-JUN-2000; 2000WO-EP05761.	
FT	PR	23-JUN-1999; 99GB-0014623.	
FT	PR	04-JAN-2001.	
FT	PR	XX	
FT	PA	(NOVS ) NOVARTIS AG.	
FT	PA	NOVARTIS-ERFINDUNGEN VERW GES MBH.	
FT	PA	XX	
FT	PI	Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;	
FT	XX		
FT	DR	WPI; 2001-137952/14.	
FT	DR	N-PSDB; AAA89353, AAA89354.	
FT	XX		
FT	PA	(NOVS ) NOVARTIS AG.	
FT	PA	NOVARTIS-ERFINDUNGEN VERW GES MBH.	
FT	PA	XX	
FT	PT	Novel gene encoding a protein that controls gene silencing, in particular silencing of plant genes	
FT	PT	Claim 1; Page 32-39; 48pp; English.	
FT	XX		
FT	CC	The present sequence is that of <i>Arabidopsis thaliana</i> protein MOM, which controls gene silencing. It is encoded by a novel silencer gene (sAA89353) of <i>A. thaliana</i> ecotype Zurich. The protein exhibits partial similarity with ATPasE/helicase proteins of the SWI2/SNF2 family. The sequence differs at amino acids 705 and 1219 compared with the protein (see AB00653), encoded by the silencer gene of wild-type <i>A. thaliana</i> ecotype Columbia. Gene silencing is useful as a molecular tool for regulating gene expression.	
FT	XX		
FT	SQ	Sequence 2001 AA;	
RESULT	4		
AB	20063		
ID	AAB20063	standard; Protein; 2001 AA.	
AC	AAB20063;		
XX			
DT	23-APR-2001	(first entry)	
XX			
DE	Arabidopsis thaliana silencing gene-encoded protein.		
KW	gene silencing; silencing gene; MOM.		
XX			
OS	Arabidopsis thaliana.		
XX			
FH	Key	Location/Qualifiers	
FT	Region	177..350	
FT	Peptide	/note= "internal repeat region"	
FT	Binding-site	362..367	
FT	Misc-difference	460..467	
FT	Region	479..719	
FT	Peptide	/note= "ATP/GTP-binding motif"	
FT	Peptide	832..838	
FT	Peptide	858..862	
FT	Domain	995..1015	
FT	Misc-difference	1219	
FT	Region	1462..1672	
FT	Region	1948..1941	
FT	Domain	1899..1941	
FT	PR	04-JAN-2001.	
FT	PR	21-JUN-2000; 2000WO-EP05761.	
FT	PR	23-JUN-1999; 99GB-0014623.	
FT	PA	(NOVS ) NOVARTIS AG.	
FT	PA	(NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
FT	PA	XX	
PI	Habu Y, Mittelsten Scheid O, Amedeo P, Paszkowski J;		





SQ	Sequence	377 AA;
FH	Key	Location/Qualifiers
FT	Region	20..211; /note= "Leucine rich repeat"
FT		
XX		
PN	WO200142285-A2.	
XX		
PD	14-JUN-2001.	
XX		
PF	05-DEC-2000; 2000WO-US32990.	
XX		
PR	10-DEC-1999; 99US-0172852.	
PR	16-DEC-1999; 99US-0172354.	
XX		
PA	(INCY-) INCYTE GENOMICS INC.	
XX		
PT	Yue H, Tang YT, Lai P, Burford N, Azimzai Y, Patterson C;	
PI	Baugh MR, Lu DAM, Shah P, Au-Young J;	
XX		
DR	WPI; 2001-381632/40.	
DR	N-PSDB; ADD08051.	
XX		
PT	New human extracellular matrix and cell adhesion molecules and polynucleotide sequences encoding them, useful for diagnosis, prevention, treatment of genetic, autoimmune and cell proliferative disorders. -	
PT		
XX		
PS	Claim 1; Page 100-101; 135pp; English.	
XX		
CC	The present sequence is a human extracellular matrix and cell adhesion molecule (XMAD). The XMAD is used for screening a compound for effectiveness as an agonist or antagonist of XMAD. The identified agonist or antagonist are used for treating a disease or condition associated with decreased or increased expression of functional XMAD. The polynucleotides encoding XMAD are useful in somatic or germline gene therapy to correct a genetic deficiency, to express a conditionally lethal gene product and to express a protein which affords protection against intracellular parasites and also for diagnosis of disorders associated with expression of XMAD. They are also used for generating hybridisation probes useful in mapping the naturally occurring genomic sequences and to create knock in humanised animals (pigs) or transgenic animals (mice or rats) to model human diseases. Oligonucleotide or longer fragments derived from the polynucleotide sequences may be used as elements on a microarray. Antibodies which specifically bind XMAD may be used for the diagnosis of disorders associated with the expression of XMAD, or in assays to monitor patients being treated with XMAD. Diseases diagnosed, prevented or treated include genetic disorders such as adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's disease, myotonic dystrophy, sickle cell anaemia, thalassaemia, autoimmune/inflammatory disorders such as acquired immune deficiency syndrome (AIDS), Addison's disease, allergies, anaemia, asthma, atherosclerosis, Crohn's disease, diabetes mellitus, atopic dermatitis, gromerulonephritis, multiple sclerosis, Grave's disease, osteoarthritis, osteoporosis, psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal, parasitic, protozoal and helminthic infections and cell proliferative disorders such as actinic keratosis, arteriosclerosis and cancer including breast, bladder, bone marrow, brain and uterus.	
CC	cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma.	
XX		
RESULT	7	
ABB5814	ID ABB5814 standard; Protein; 1284 AA.	
XX		
DE	Drosophila melanogaster polypeptide SEQ ID NO 3234.	
XX		
KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.	
XX		
OS	Drosophila melanogaster.	
XX		
PN	WO200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	23-MAR-2001; 2001WO-US09231.	
XX		
PR	23-MAR-2000; 2000US-191637P.	
PR	11-JUL-2000; 2000US-0614150.	
XX		
PA	(PERE ) PE CORP NY.	
XX		
PI	Venter JC, Adams M, Li PHD, Myers EW;	
XX		
DR	WPI; 2001-656800/75.	
DR	N-PSDB; ABL02917.	
XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -	
PT		
PS	Disclosure; SEQ ID NO 3234; 21pp + Sequence Listing; English.	
XX		
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL1676-ABL1675), expressed DNA sequences (ABL1640-ABL16175) and the encoded proteins (ABB5773-ABB2072).	
CC		

xx  
PS  
xx  
Disclosure; Fig 2; 55pp; English

XX	AC	AAR15160;
XX	AC	
DT	AC	
XX	AC	17-DEC-2001 (updated)
DT	AC	28-FEB-1992 (first entry)
DE	AC	FB29 chicken c-ski protein.
XX	AC	
KW	AC	Transgenic; muscle growth; muscle d-
XX	AC	
OS	AC	Gallus domesticus.
XX	PN	
XX	PN	USN7620415-N.
PD	PN	
XX	PN	05-NOV-1991.
PP	PN	
XX	PN	03-DEC-1990; 900US-0620415.
PR	PN	
XX	PR	03-DEC-1990; 900US-0620415.
PA	PR	
XX	PA	(USSH ) NAT INST OF HEALTH.
PI	PA	
Hughes S, Sutrave P;	PI	
DR	PI	
DR	PI	WPI: 1991-361684/49.
DR	PI	N-PSDB, AAQ14799.
XX	PT	
XX	PT	DNA encoding chicken C-ski proteins
PT	PT	animals having increased muscle grow-
degeneration or obesity	PT	

generation; obesity.

<b>Domain</b> <b>Key Domain</b> <b>Mut sp.</b> <b>Domain</b>	<b>T</b> 9 <b>9994</b> <b>ARR96994</b> strand <b>AAR96994;</b> <b>16-AUG-1996</b> ( <b>Mouse IRS-2.</b> <b>IRS-2; insulin</b> <b>diagnosis; vec-</b> <b>mus sp.</b>	<b>231</b> <i>rannunculus</i> <b>395</b> <i>safflowerspab-</i> <b>282</b> <i>VSVFORSSSE</i> <b>455</b> <i>atvvspssse</i> <b>1;</b> <b>1;</b> <b>1;</b> <b>1;</b>
---	---	---

PPKMKVSPPPQQPQAVARTV  
vsanekeishthipalir  
|| || : || |  
| | : || |  
PPPSAAQPKRKAETT

DSSG-  
|||  
dssfsy  
307  
480  
1 AA.

Sppk-syksfest

NASTP NAVAP

nvalapPPG

प्राचीन  
कव्य

ISPEPPRC  
Suppo  
vws

: 454  
, 281

FT	Domain	481..527	CC receptors. It contains a number of phosphorylation sites in common with IRS-1, showing the 2 proteins to be related.
FT	/label= IH3	/note= "IRS-homology domain III"	IRS-2
FT	74..80	/note= "common phosphorylation site"	can be obtid. e.g. by affinity purification from insulin-stimulated
FT	Modified-site	/note= "unique phosphorylation site"	FDC-B2 cells using immobilized SH2 domains of p85, or can be expressed in transformed host cells carrying an encoding cDNA sequence (ATR28293).
FT	120..125	/note= "unique phosphorylation site"	Cells or animals having the IRS-2 transgene can be used to study insulin-related disorders, e.g. type II diabetes. IRS-2 can be used to raise antibodies and for drug screening.
FT	135..140	/note= "common phosphorylation site"	CC
FT	180..185	/note= "unique phosphorylation site"	CC
FT	Modified-site	/note= "common phosphorylation site"	CC
FT	499..504	/note= "unique phosphorylation site"	XX
FT	Modified-site	/note= "common phosphorylation site"	SQ Sequence 1321 AA:
FT	535..540	/note= "common phosphorylation site"	Query Match 6.0%; Score 103.5; DB 17; Length 1321;
FT	593..598	/note= "common phosphorylation site"	Best Local Similarity 22.2%; Pred. No. 1.3; Matches 73; Conservative 30; Mismatches 93; Indels 133; Gaps 16; Matches 73; Conservative 30; Mismatches 93; Indels 133; Gaps 16;
FT	623..628	/note= "common phosphorylation site"	Db 806 scs-----gndqyv1musp--vgrileeerie-----pgrtpgagt----- 841
FT	627..632	/note= "common phosphorylation site"	Qy 90 RGSKLPLRSKKGVHVYRELSPRTSGQSQRKVFDHVSYTMQOLESTINEGEPFHAGDRP 149
FT	648..653	/note= "unique phosphorylation site"	Db 842 -----gaqshsq---phsavps-----snr-----psaligrp 869
FT	670..675	/note= "common phosphorylation site"	Qy 150 SKHLKLVLTRHCLRAIRLPRISID-----LMESPLNLSGEALSPATAKDELTOML 202
FT	Modified-site	/note= "common phosphorylation site"	Db 870 eg-----flgqrscravprtrsliegltipsneyplpt---epksp----- 908
FT	733..738	/note= "unique phosphorylation site"	Qy 203 KSAARSELGMWYSKRCFYLRRARRRKFAMWPKVLOISMKPYMERITPMAYRQSGPP 262
FT	757..762	/note= "common phosphorylation site"	Db 909 -----geyin----ldfgaqtrisppapilasaasssisasspass19sgtpg 956
FT	813..818	/note= "unique phosphorylation site"	Qy 263 KNA-----spslpgpkn-----isprsvysvorssrpkn 294
FT	899..903	/note= "unique phosphorylation site"	Db 957 tsdsrqgrspidsymntdsspkspkpkstrsqtvsmdgl1lsp--eassppplprps 1014
FT	910..915	/note= "common phosphorylation site"	Qy 295 VSP-----PQQPARVARAKSKVSAASO 316
FT	969..974	/note= "common phosphorylation site"	Db 1015 tpsqlsqqlppapgdlyrippasaatsq 1043
FT	Modified-site	/note= "unique phosphorylation site"	RESULT 10
FT	1241..1246	/note= "unique phosphorylation site"	AAB23012
FT	1302..1307	/note= "common phosphorylation site"	ID AAB23012 standard; Protein: 2742 AA.
FT	Modified-site	/note= "common phosphorylation site"	XX
FT	1302..1307	/note= "common phosphorylation site"	AC AAB23012;
XX	PN W09610629-A1.	XX	XX DT 16-JAN-2001 (first entry)
XX	PN 11-APR-1996.	XX	XX DE Human APC protein (splice variant 2).
XX	PN 03-OCT-1995;	XX	XX APC gene; Adenomatous Polyposis Coli gene; human; chromosome 5q21;
PF	95W0-US13041.	PF	KW familial adenomatous polyposis; FAP locus; Gardner's syndrome; GS;
XX	03-OCT-1994;	PR 94US-0317310.	KW sporadic tumour; adenoma; carcinoma; cancer; lung; breast; colon; rectum;
XX	(JOSL-) JOSLIN DIABETES CENT INC.	PR	KW bladder; liver; sarcoma; prostate; leukaemia; lymphoma;
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	PA	KW tumour suppressor; anti-APC antibody; detection; diagnosis; prognosis;
XX	Pierce JH, Sun XJ, White MF;	XX	KW genetic predisposition; drug screening; DP2.5; splice variant.
XX	PN W09610629-A1.	OS Homo sapiens.	XX PN US6114124-A.
DR	WDI; 1996-209351/21.	XX	XX PD 05-SEP-2000.
DR	N-PSDB; AAT28293.	XX	XX PF 25-MAY-1995; 95US-0450592.
PT	New insulin receptor substrate polypeptide and corresp. nucleic acid	XX	XX PR 16-JAN-1991; 91GB-0000942.
PT	- vectors, antibodies etc., useful for diagnosis, treatment and	XX	XX PR 16-JAN-1991; 91GB-0000943.
PT	assessing risk of diabetes etc., also for drug screening	XX	XX PR 16-JAN-1991; 91GB-0000944.
XX	Disclosure; Page 47-53; 105pp; English.	XX PR 16-JAN-1991; 91GB-0000945.	
CC	Mouse insulin receptor substrate-2, or IRS-2 (AAR96994), is a	XX PR 01-AUG-1991; 91US-0741940.	
CC	substrate for the insulin receptor, interleukin-4 receptor and	XX PR 12-AUG-1994; 94US-0289548.	
CC	interleukin-15 receptor, and can be phosphorylated by these		

ID AAR23962 standard; Protein; 2783 AA.  
 XX  
 AC AAR23962;  
 XX  
 DT 06-JAN-1993 ( first entry )  
 DE AFP-1.  
 XX  
 OS Homo sapiens .  
 XX  
 FH Enhancer; alpha fetoprotein; homeodomain; zinc-finger.  
 XX  
 Key  
 FT Domain 1231..1291  
 FT /note= "homeodomain"  
 FT Domain 1308..1388  
 FT /label= homeodomain  
 FT Domain 1727..1786  
 FT /label= homeodomain  
 FT Domain 2032..2091  
 FT /label= homeodomain  
 FT Domain 67..99  
 FT /label= zinc-finger  
 FT Domain 123..155  
 FT /label= zinc-finger  
 FT Domain 171..203  
 FT /label= zinc-finger  
 FT Domain 306..366  
 FT /label= zinc-finger  
 FT Domain 443..476  
 FT /label= zinc-finger  
 FT Domain 484..543  
 FT /label= zinc-finger  
 FT Domain 628..660  
 FT /label= zinc-finger  
 FT Domain 679..711  
 FT /label= zinc-finger  
 FT Domain 1066..1097  
 FT /label= zinc-finger  
 FT Domain 1411..1442  
 FT /label= zinc-finger  
 FT Domain 1613..1643  
 FT /label= zinc-finger  
 FT Domain 1790..1829  
 FT /label= zinc-finger  
 FT Domain 2107..2139  
 FT /label= zinc-finger  
 FT Domain 2540..2571  
 FT /label= zinc-finger  
 FT Domain 2606..2638  
 FT /label= zinc-finger  
 XX EP487229-A.  
 PN  
 XX  
 PR 07-NOV-1991; 91EP-0310334.  
 XX  
 PR 07-NOV-1990; 90JP-0301412.  
 XX  
 PA (SNOW ) SNOW BRAND MILK PROD CO LTD.  
 PI Higashio K, Morinaga T, Tamaki T, Yasuda N,  
 XX WPI; 1992-176828/22.  
 DR N-PSDB; AAQ24828.  
 XX  
 PT DNA encoding protein binding to alpha-fetoprotein gene enhancer -  
 PT useful for prodn. of biological active protein  
 XX



KW screening; prostate cancer; benign prostatic hyperplasia;  
 KW hair loss; alopecia; therapy.  
 XX OS Homo sapiens.  
 XX PN WO20017332-A1.  
 XX PD 04-OCT-2001.  
 XX PF 26-MAR-2001; 2001WO-US09378.  
 XX PR 24-MAR-2000; 2000US191768P.  
 XX PR 15-AUG-2000; 2000US-225618P.  
 XX PA (UYNY ) UNIV NEW YORK STATE.  
 XX PT Garabedian M, Taneja S, Hittelman A, Markus S;  
 XX DR WPI; 2001-61152/70.  
 XX N-PSDB; AAH26896.  
 XX PT New androgen receptor transcriptional coregulatory proteins that  
 PT interact with androgen receptor to regulate androgen-dependent gene  
 expression, useful for producing antibodies to treat androgen dependent  
 diseases.  
 XX PS Claim 31; Page 99-107; 113pp; English.  
 CC The present sequence is that of a novel human androgen receptor  
 CC trapped protein (ART), which belongs to a new class of proteins  
 CC that interact with the N-terminus of the androgen receptor as  
 CC androgen receptor transcriptional coregulator (i.e., coactivator)  
 CC proteins, regulating androgen-dependent gene expression. To identify  
 CC proteins that interact with the androgen receptor N-terminus,  
 CC a modified yeast two-hybrid system that allowed the identification  
 CC of factors expressed in the prostate and which associated with  
 CC transcriptional activators was used. An androgen-stimulated LNCaP  
 CC prostate cancer cell cDNA library fused to the LexA DNA binding  
 CC domain was screened for proteins that interacted with the androgen  
 CC receptor N-terminal transcriptional activation domain (amino acid  
 residues 18-500). From approximately 1 million library  
 CC transforms, 8 clones were isolated, and termed ARTs, for  
 CC androgen receptor trapped. ARTs are useful for producing  
 CC antibodies (or molecules having the binding portion of an antibody)  
 CC which can be administered to patients to treat androgen dependent  
 CC diseases, especially prostate cancer, benign prostatic hyperplasia  
 CC or androgen-dependent hair loss. They can be used to identify  
 CC inhibitors disrupting interaction between an androgen receptor and  
 CC an androgen receptor transcriptional coregulatory protein, useful  
 XX to inhibit such interaction (all claimed) therapeutically.  
 SQ Sequence 2783 AA;

Query Match 5.9%; Score 103; DB 22; Length 2783;  
 Best Local Similarity 28.0%; Pred. No. 4.1; Mismatches 44; Indels 44; Gaps 8;

Oy 171 SIDLME-----SPLPNLGEALSPTA-TAKDITONILKSAARSELGMVSKRQ 218  
 Db 1454 Smdamilmptssscstpmcsayapsapansantassafql---taeeklatfnke 1510  
 Oy 219 EFLYLRARRRRRFKFW-----KPVLOISEMKPKMVFHFMAYRDGSPP 262  
 Db 1511 ---agdekpklaeaepsaqpnqtqekqqgkpbqlq--qgeqeqdktntpqq----klp 1558  
 Oy 263 KNASTPSLPEPKNTPRQVSPRQSSPPKKNVSPP 299  
 Db 1559 qivs1pslppqapqpcipq-sspspsqishlp 1594

RESULT 14  
 AR83508

---

ID IAR63508 standard; Protein; 2842 AA.  
 XX AC AAR63508;  
 XX DT 23-MAY-1995 (first entry)  
 XX DE Adenomatous polyposis coli tumour repressor.  
 XX KW Adenomatous polyposis coli; tumour repressor; Gardner's syndrome;  
 KW familial adenomatous polyposis; cancer diagnosis and prognosis;  
 KW tumorigenesis suppression.  
 XX OS Homo sapiens.  
 XX PN US5352775-A.  
 XX PD 04-OCT-1994.  
 XX PF 08-AUG-1991; 91US-0741940.  
 XX PR 16-JUN-1991; 91GB-0000962.  
 PR 16-JUN-1991; 91GB-0000963.  
 PR 16-JAN-1991; 91GB-0000974.  
 PR 16-JAN-1991; 91GB-0000975.  
 XX PA (CANC-) CANCER INST.  
 PA (ICIL) IMPERIAL CHEM IND.  
 PA (UYJO) UNIV. JOHNS HOPKINS.  
 PA (UTAH) UNIV UTAH.  
 XX DR N-PSDB; AAQ72297.  
 XX PT New human adenomatous polyposis coli DNA encoding tumour  
 PT repressor - and derived primers and probes for diagnosis,  
 XX prognosis and treatment of cancer.  
 XX PS Claim 3; Columns 71-84; 113pp; English.  
 XX AAQ72297 is a cDNA isolated from the human adenomatous polyposis  
 CC coli (APC) gene, it encodes the tumour repressors described in AAR63507  
 CC and AAR63508. Determination of alterations in APC or its expression  
 CC products, can be used for the diagnosis and prognosis of cancer e.g.  
 CC colorectal, lung and breast tumours; and for determining predisposition  
 CC to certain cancers such as familial adenomatous polyposis (FAP) and  
 Gardner's syndrome. The wild type APC gene (or a part of it) can be used  
 CC therapeutically to restore gene function, while primers and probes  
 CC derived from the cDNA (AAQ72297, AAQ72298 and AAQ7241-568) can be used to  
 CC detect mutations. Also APC proteins or analogues can be administered to  
 CC compensate for a defective gene and epithelial cells, or transgenic  
 CC animals carrying a mutated APC allele are useful for detecting  
 CC therapeutic agents able to suppress tumorigenesis.  
 XX SQ Sequence 2842 AA;

Query Match 5.9%; Score 103; DB 15; Length 2842;  
 Best Local Similarity 22.3%; Pred. No. 4.2; Mismatches 57; Indels 46; Gaps 11;

Oy 79 PFLEROPKKWHRGSKLPRFSKHGVHVRERSPHGQSOSREVDFHSYMMQOLESTIRNE 138  
 Db 2151 pdacekptfskpkpilpkgektstkkeseekgkgkv-ykslitgkvs--- 2204  
 Oy 139 GPEPHLAGDRPSKHLKLVFIRCHLRLRIPRISIDLMSPNL-SGEALSPTATAKE 196  
 Db 2205 -nseisq-----mkqp1qa-nmuisrgrtmihlgvnrssstspsvskgp 2251  
 Oy 197 ITOMILKSAARSELGMVSKREFYIIRRARRRFKAWKPVLOISEMKPKMVFHFMAYR 256

```

Db 2252 lktptkspkspseqgat-tspr-----gakpsv--seispar-qtsqgg 229
  :|| :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|
Qy 257 DSQSPPKNA--SISLSPGPKNISPPRQVSVQSSPPKNNISPPQPAVARASKYSA 313
  :|| :| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|
Db 2294 sskkapssgsksdtspsrpagqplsrpinqspgnsispgngipspnqklsqprtspta 235
  :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|
Qy 314 ASQQVQRNRGKSLY 329
  :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|
Db 2354 st---ksqsg9kmsy 2365
  :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|

```

RESULT 15

AAR26052	ID	AAR26052	standard; Protein;	2843 AA.
XX	XX	XX		
AC	AAR26052;			
XX	DAT	28-JAN-1993	(first entry)	
XX	DE	APC gene product in familial adenomatous polyposis.		
XX	KW	neoplasm; cancer; oncogene; tumour; growth; defecation; diagnosis; prognosis; treatment; sporadic colorectal carcinomas; ss.		
XX	OS	Homo sapiens.		
XX	PN	W09213103-A.		
XX	PD	06-AUG-1992.		
XX	PF	92WO-US00376.		
XX	PR	16-JAN-1991; 91GB-0000963.		
PR	08-AUG-1991; 91US-0741940.			
XX	PA	(CANC-) CANCER INST.		
PA	(ICIL ) IMPERIAL CHEM IND PLC.			
PA	(UYTO ) UNIV JOHNS HOPKINS.			
PA	(UTAH ) UNIV UTAH.			
XX	PI	Albertsen H, Aviand R, Carlson ML, Groden JL, Hedge PJ, Thliveris A;		
PI	Joslyn' Kinzler KW, Markham A, Nakamura Y, Vogelstein B, Whitecll, Markham AF;			
XX	DR	WPI; 1992-284685/34.		
XX	N-PSB;	AAQ27234.		
PT	Detection of somatic and germ-line alterations of human APC gene - used to diagnose, treat and study familial adenomatous polyposis and sporadic colorectal cancer			
PT				
XX	PS	Disclosure; Page 47; 132pp; English.		
XX	CC	This sequence is encoded by the APC (Adenomatous Polyposis Coli) gene associated with tumorigenesis, found on chromosome 5q. The sequence may be mutated by deletions, insertions, inversions, or point mutations of the gene. The APC gene is expressed in most normal tissues as well, suggesting that APC is a tumour suppressor.		
XX	SQ	Sequence 2843 AA;		

Query Match 5.9%; Score 103; DB 13; Length 2843;  
 Best Local Similarity 22.3%; Pred. No. 4 2; Mismatches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11; Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

Qy 79 PFLURQPKVWHGSKLPLPRKQVHVRLELRPRTGQSQRKVFDVSYVTHQOLASIRNE 138
 :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|
Db 2152 pdqgeektsnkgprlkpkgeekstetkkieskqgkky--yksiltlgks--- 2205
 :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|
Qy 139 GPEPHLADPRSKHLKLVFRHCIRALLPRSIDLEMSPPLNL--SGEALSPTAKDE 196
 :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|

```

Db 2206 --nseisqg-----nkqlqa-nmpssrgtmihiavgrrsssstpuskkgpp 2252
Qy 197 itomilksaarselgmyvskrqeylrrarrkrkawkvlylosisebmkkpmefhtpmayr 256
Db 2253 lktpaskpsgeqtat-tspr-----gakpsvk--selspvar-qtsqgg 2294
Qy 257 DSGSPKNA--STPSLPGPKNISPPRQVSPQRQSPSPKKNISPPQPAVARYASKVSA 313
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 2295 askapsrsrsdpsrpaqgqisprqsgprnsispgngispnklsqiptuspsata 2354
Qy 314 ASQQVQRNRGNAKSLTY 329
   :: : : | : | : |
Db 2355 st---ksssgkmsy 2366

```





RESULT 4  
US-08-289-548A-7  
; Sequence 7, Application US/08289548A  
; General Information:  
; Applicant: KINZLER, KENNETH  
; Applicant: MARKHAM, ALEXANDER F.  
; Applicant: NAKAMURA, YUSKE  
; Applicant: CARLSON, MARY  
; Applicant: GRODEN, JOANNA  
; Applicant: HEDGE, PHILIP J.  
; Applicant: JOSLYN, GEOFF  
; Applicant: KINZLER, KENNETH  
; Applicant: MARKHAM, ALEXANDER F.  
; Applicant: NAKAMURA, YUSKE  
; Applicant: CARLSON, MARY  
; Applicant: GRODEN, JOANNA  
; Applicant: HEDGE, PHILIP J.  
; Applicant: JOSLYN, GEOFF  
; Correspondence Address:  
; Addressee: Banner & Allegretti, LTD  
; Street: 1001 G Street, NW  
; City: Washington  
; State: D.C.  
; Country: USA  
; ZIP: 20001-4598  
; Computer Readable Form:  
; Medium Type: Floppy disk  
; Computer: IBM PC compatible  
; Operating System: PC-DOS/MS-DOS  
; Software: Patentin Release #1.0, Version #1.25  
; Current Application Data:  
; Application Number: US/08/289,548A  
; Filing Date: 12-AUG-1994  
; Classification: 435  
; Attorney/Agent Information:  
; Name: Kagan, Sarah A.  
; Registration Number: 32,141  
; Reference/Docket Number: 1107.46943  
; Telecommunication Information:  
; Telephone: 202-508-9100  
; Telefax: 202-508-9299  
; Information for Seq ID No: 7:  
; Sequence Characteristics:  
; Length: 2842 amino acids  
; Type: amino acid  
; Strandedness: single  
; Topology: linear  
; Molecule Type: protein  
; Original Source: Homo sapiens  
; Immediate Source: APC  
; Clone: APC

RESULT 5  
US-08-452-654-7  
; Sequence 7, Application US/08452654  
; General Information:  
; Applicant: ALBERGEN, HANS  
; Applicant: ANAND, RAKESH  
; Applicant: CARLSON, MARY  
; Applicant: GRODEN, JOANNA  
; Applicant: HEDGE, PHILIP J.  
; Applicant: JOSLYN, GEOFF  
; Applicant: KINZLER, KENNETH  
; Applicant: MARKHAM, ALEXANDER F.  
; Applicant: NAKAMURA, YUSKE  
; Applicant: THILIVERIS, ANDREW  
; Title of Invention: INHERITED AND SOMATIC MUTATIONS OF APC  
; Number of Sequences: 102  
; Correspondence Address:  
; Addressee: Banner, Birch, McKie & Beckett  
; Street: 1001 G Street, NW  
; City: Washington  
; State: D.C.  
; Country: USA  
; ZIP: 20001-4598  
; Computer Readable Form:  
; Medium Type: Floppy disk  
; Computer: IBM PC compatible  
; Operating System: PC-DOS/MS-DOS  
; Software: Patentin Release #1.0, Version #1.25  
; Current Application Data:  
; Application Number: US/08/452,654  
; Filing Date: 25-MAY-1995  
; Classification: 536  
; Prior Application Data:  
; Application Number: US 07/741,940  
; Filing Date: 08-AUG-1991  
; Attorney/Agent Information:  
; Name: Kagan, Sarah A.  
; Registration Number: 32,141  
; Reference/Docket Number: 1107.035574  
; Telecommunication Information:  
; Telephone: 202-508-9100  
; Telefax: 202-508-9299  
; Information for Seq ID No: 7:  
; Sequence Characteristics:  
; Length: 2842 amino acids  
; Type: amino acid  
; Strandedness: single  
; Topology: linear  
; Molecule Type: protein

Query Match 5.9%; Score 103; DB 1; Length 2842;  
Best Local Similarity 22.3%; Pred. No. 0.49;  
Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-452-654-7

ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: APC  
US-08-452-654-7

Query Match 5.9%; Score 103; DB 1; Length 2842;

Best Local Similarity 22.3%; Pred. No. 0.49; Indels 46; Gaps 11; Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

Qy 79 PFLERQPKVHRSKILPRESKHWYHWRLESPHTGQSOSRKFDHSVMMQLESIRNE 138  
Db 2151 PDQEERPKFTSNKGPRILKPGSKSTLETKRKESEKGIGKGVY-KSLITGVRS--- 2204

Qy 139 GPEPHLAGRPSRKHKLVLFRHCLRLRPRISIDLMESPLNL-SGEALSPVATADE 196  
Db 2205 -NSEISGQ-----MKQPLQA-NMPSISRGRTMHIGPVNRNSSTSIVSKKGPP 2251

Qy 197 ITOMILKSARSELGMVSKRQEFYLARRRKFAWKVLOISEKPKWNEFHTPMAYR 256  
Db 2252 :LTPRKSSEGOTAT-TSPR-----GAKSVK--SESPVYR-OTSQIG 2293

Qy 257 DSGSPKNA--STPSLPGPKNISPROVSYQRSSPPKNSPPROPAVARYSKA 313  
Db 2294 SSKASRSRSRSDTSPRSPAQQLSPRQSPGPNSTSPGNISPPNKLSQLPRISPTA 2353

Qy 314 ASQVQVRNGRGNASLY 329  
Db 2354 ST-----KSSGSKMSY 2365

RESULT 6  
US-07-741-940-2

Patent No. 5352775  
GENERAL INFORMATION:  
APPLICANT: ALBERSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: TULIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESS: Bone, Birch, McKie & Beckett

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/741,940

FILING DATE: 1992/10/09

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,441

REFERENCE/DOCKET NUMBER: 1107.035574

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2843 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULAR TYPE: protein

RESULT 7  
US-08-289-508A-2

Sequence 2, Application US/08289548A

Patent No. 5648212

GENERAL INFORMATION:

APPLICANT: ALBERSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: TULIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/289,548A

FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.46943

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9299  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-289-548A-2

Query Match 5.9%; Score 103; DB 1; Length 2843;  
 Best Local Similarity 22.3%; Pred. No. 0.49; Mismatches 57; Conservatism 46; Indels 46; Gaps 11; Matches 57;

QY 79 PFLEROPKVVRHGRSLRPLRSKGVHIVRLRSPTPHGSQOSRKVFDYHVSMTMQLESIRNE 138  
 Db 2152 PDQEKEPKFTSNKGPRILPKGEKSTLETKIESKIKGGKV-YKSLITGKVS--- 2205

QY 139 GPEPHLAGDRPSKHLKLVLTRICRALKRLRPTSIDLMESPPLPNL--SGEALPTAAKDE 196  
 Db 2206 --NSEISQ-----MKQPLQA-NMPSISGRIMIHPGVYRNSSSITSPVKGPP 2252

QY 197 ITOMILKSAARSELGMYVKROFYLRARRRKFAWKPVQIQSSEMKVFMFHTPMAYR 256  
 Db 2253 LKTPASKPSSEGOTAT-TPSR-----GAKPSVK--SELSEVAR-OPSQIGG 2294

QY 257 DSGGPKNA--STSPLDPKNSIPRQVSVPORSSPPKNSPPQPAVARTRSKYA 313  
 Db 2295 SSKAAPSRSRSRSTPSRPAQQLSPRPIQSPQRNSISPGRNGISPNNKLSQLPRTSSPSTA 2354

QY 314 ASQVQNRNGMAKSY 329  
 Db 2355 ST---KSSGSKMSY 2366

RESULT 8  
 US-08-452-654-2  
 Sequence 2, Application US/08452654  
 GENERAL INFORMATION:  
 PATENT NO. 569154

APPLICANT: ALBERTSEN, HANS  
 APPLICANT: ANAND, RAKESH  
 APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JOANNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: JOSLIN, GEOFF  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THLIVERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 NUMBER OF SEQUENCES: 94  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner, Birch, McKie & Beckett  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/452,654  
 FILING DATE: 25-MAY-1995  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/741,940  
 FILING DATE: 08-AUG-1991

Query Match 5.9%; Score 103; DB 1; Length 2843;  
 Best Local Similarity 22.3%; Pred. No. 0.49; Mismatches 57; Conservatism 46; Indels 4; Gaps 4; Matches 57;

QY 79 PFLEROPKVVRHGRSLRPLRSKGVHIVRLRSPTPHGSQOSRKVFDYHVSMTMQLESIRNE 138  
 Db 2152 PDQEKEPKFTSNKGPRILPKGEKSTLETKIESKIKGGKV-YKSLITGKVS--- 2205

QY 139 GPEPHLAGDRPSKHLKLVLTRICRALKRLRPTSIDLMESPPLPNL--SGEALPTAAKDE 196  
 Db 2206 --NSEISQ-----MKQPLQA-NMPSISGRIMIHPGVYRNSSSITSPVKGPP 2252

QY 197 ITOMILKSAARSELGMYVKROFYLRARRRKFAWKPVQIQSSEMKVFMFHTPMAYR 256  
 Db 2253 LKTPASKPSSEGOTAT-TPSR-----GAKPSVK--SELSEVAR-OPSQIGG 2294

QY 257 DSGSPKNA--STSPLDPKNSIPRQVSVPORSSPPKNSPPQPAVARTRSKYA 313  
 Db 2295 SSKAAPSRSRSRSTPSRPAQQLSPRPIQSPQRNSISPGRNGISPNNKLSQLPRTSSPSTA 2354

QY 314 ASQVQNRNGMAKSY 329  
 Db 2355 ST---KSSGSKMSY 2366

RESULT 9  
 US-08-452-655B-2  
 Sequence 2, Application US/08452655B  
 PATENT NO. 5783666  
 GENERAL INFORMATION:  
 APPLICANT: ALBERTSEN, HANS  
 APPLICANT: ANAND, RAKESH  
 APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JOANNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: JOSLIN, GEOFF  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 APPLICANT: THLIVERIS, ANDREW  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 NUMBER OF SEQUENCES: 102  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner & Witcoff, Ltd.  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/452,655B  
 FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:  
 NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107-035574  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELERFAX: 202-508-9399  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-452-655B-2

NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107-035574  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELERFAX: 202-508-9399  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-452-655B-2



ADDRESSEE: BANNER & WITCOFF, LTD.  
 STREET: 1001 G STREET, N.W.  
 CITY: WASHINGTON  
 STATE: DC  
 COUNTRY: US  
 ZIP: 20001

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/370,235A  
 FILING DATE: 01-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:

NAME: KAGAN, SARAH A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 01107.48688

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202 508 9100  
 TELEFAX: 202 508 9299

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

US-08-370-235A-2

Query Match 5.9%; Score 103; DB 2; Length 2843;  
 Best Local Similarity 22.3%; Pred. No. 0.49; Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PFLERQPKVHRSKSLPRFSKGVHVRRLSPTHQSQRKVFDIHSYTMQLESIRNE 138  
 Db 2152 PDQESEKPTSNKGPRLKPGKSKTLETKIESKGKGVV--YKLITGKVS--- 2205

QY 139 GPEPHLAGDRPSKHLKLVLFRHCLRALRUPRTSIDLMESPLPNL--SGEASPTATAKDE 196  
 Db 2206 --NSEISGO-----MKQPLQA-NMPSISRGTMHIIPGYRNSSSSSPVSKGP 2352

QY 197 ITQMLKSAARSELGMVSKRQEFTYLRRRKFAWKPVQLQSSEMKPVMFHTPMAYR 256  
 Db 2253 LKTPASKSPSKEGOTAT-TSR-----GAKPSVK--SELSPVAR-QTSQIG 2294

QY 257 DSGSPKNA--STPSLPGKPNISPPROVSYFQRSSPPKNSPPQPAFVARTASKYSA 313  
 Db 2295 SSKAPSRSGSDTSPSRPAQPLSRPQPLSPGRNNSISPPGRNGISPNNKLSQPLRTSPSTA 2354

QY 314 ASQQVQRNRGNNAKSLY 329  
 Db 2355 ST---KSSGSKMSY 2366

RESULT 12

US-08-450-582-2

; Sequence 2, Application US/08450582

; Patent No. 6114124

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS  
 ; APPLICANT: ANAND, RAKESH  
 ; APPLICANT: CARLSON, MARY  
 ; APPLICANT: GRODEN, JOANNA  
 ; APPLICANT: HEDGE, PHILIP J.  
 ; APPLICANT: JOSLYN, GEOFF  
 ; APPLICANT: KINZLER, KENNETH  
 ; APPLICANT: MARKHAM, ALEXANDER F.  
 ; APPLICANT: NAKAMURA, YUSUKE  
 ; APPLICANT: THIETERS, ANDREW  
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner & Witcoff, Ltd.  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/450,582

FILING DATE:

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/741,940  
 FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107.49964

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

US-08-450-582-2

Query Match 5.9%; Score 103; DB 3; Length 2843;  
 Best Local Similarity 22.3%; Pred. No. 0.49; Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PFLERQPKVHRSKSLPRFSKGVHVRRLSPTHQSQRKVFDIHSYTMQLESIRNE 138  
 Db 2152 PDQESEKPTSNKGPRLKPGKSKTLETKIESKGKGVV--YKLITGKVS--- 2205

QY 139 GPEPHLAGDRPSKHLKLVLFRHCLRALRUPRTSIDLMESPLPNL--SGEASPTATAKDE 196  
 Db 2206 --NSEISGO-----MKQPLQA-NMPSISRGTMHIIPGYRNSSSSSPVSKGP 2352

QY 197 ITQMLKSAARSELGMVSKRQEFTYLRRRKFAWKPVQLQSSEMKPVMFHTPMAYR 256  
 Db 2253 LKTPASKSPSKEGOTAT-TSR-----GAKPSVK--SELSPVAR-QTSQIG 2294

QY 257 DSGSPKNA--STPSLPGKPNISPPROVSYFQRSSPPKNSPPQPAFVARTASKYSA 313  
 Db 2295 SSKAPSRSGSDTSPSRPAQPLSRPQPLSPGRNNSISPPGRNGISPNNKLSQPLRTSPSTA 2354

QY 314 ASQQVQRNRGNNAKSLY 329  
 Db 2355 ST---KSSGSKMSY 2366

RESULT 13

US-08-450-582-7

; Sequence 7, Application US/08450582

; Patent No. 6114124

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS  
 ; APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY  
 APPLICANT: GRODEN, JONNA  
 APPLICANT: HEDGE, PHILIP J.  
 APPLICANT: KINZLER, KENNETH  
 APPLICANT: MARKHAM, ALEXANDER F.  
 APPLICANT: NAKAMURA, YUSUKE  
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 NUMBER OF SEQUENCES: 102  
 CORRESPONDENCE ADDRESS:  
 STREET: 1001 G Street, NW  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20001-4598

COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/450,582

CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/452,655  
 FILING DATE: 25-MAY-1995  
 APPLICATION NUMBER: US 08/289,548  
 PRIOR APPLICATION DATA:  
 FILING DATE: 12-AUG-1994

APPLICATION NUMBER: US 07/741,940  
 FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:  
 NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,141  
 REFERENCE/DOCKET NUMBER: 1107.49964

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2843 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ANTI-SENSE: NO  
 US-08-450-582-7

Query Match 5.9%; Score 103; DB 3; length 2843;  
 Best Local Similarity 22.3%; Pred. No. 0.49;  
 Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PPLEROPKVVHRGSKILREFSKHGVHVRRLSPTHGSQQSKRVDIHSYTMWOLESTRNE 138  
 Db 2152 PDEQEPRFSNKPRPIKPGKSTLETKKIESSGKGKVV--YKSUTGKVRS--- 2205

QY 139 GPEPHLAGDGPDKPSHLKLVFIHCRLRUPRISIDMESPLNL--SSEALSPATAKDE 196  
 Db 2206 --NSEISGQ-----MKPQLA--NMPISGRGMHIIHGVRNSSSTSVSKKGPP 2252

QY 197 ITQMLIKSAARSELGMMVVKSQEFLYARRRARRRKFAWKPVQIQTSEMKEVMEHTPMAYR 256  
 Db 2253 LKTPASKSSSEGQAT-TSPR-----GAKPSVK--SELSVAR-QTSQIGG 2294

QY 257 DSGSPPKNA--SPTSPGKPNISPROVSPVSPORSSPPQPAFPARTASKYSA 313  
 Db 2295 SSKASRSRSRSDTSPSRQPLSPRIOSPPGRNISGRRNGISPPNKLSQLRTSPSTA 2354

---

QY 314 ASQOVNRQNAKSLY 329  
 Db 2355 ST-----KSSSGKMSY 2366

RESULT 14  
 ; Sequence 7, Application US/08821355A  
 ; Patient No. 585175  
 GENERAL INFORMATION:  
 APPLICANT: Barker, Nick  
 APPLICANT: Clevers, Hans  
 APPLICANT: Korinek, Vladimir  
 APPLICANT: Morin, Patrice  
 APPLICANT: Kinzler, Kenneth  
 APPLICANT: Vogelstein, Bert  
 APPLICANT: Sparks, Andrew  
 PRIORITY: Beta Catenin, TCF-4, and APC  
 TITLE OF INVENTION: Interact to Prevent Cancer  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Banner & Witcoff, Ltd.  
 STREET: 1001 G Street, N.W.  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20001

COMPUTER READABLE FORM:  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/821,355A  
 FILING DATE: 20-MAR-1997  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kagan, Sarah A.  
 REGISTRATION NUMBER: 32,145  
 REFERENCE/DOCKET NUMBER: 1107.05064

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-508-9100  
 TELEX: 97430 BMB UT  
 TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2973 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: No. 585175e  
 US-08-821-355A-7

Query Match 5.9%; Score 103; DB 2; length 2973;  
 Best Local Similarity 22.3%; Pred. No. 0.52;  
 Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PPLEROPKVVHRGSKILREFSKHGVHVRRLSPTHGSQQSKRVDIHSYTMWOLESTRNE 138  
 Db 2152 PDEQEPRFSNKPRPIKPGKSTLETKKIESSGKGKVV--YKSUTGKVRS--- 2205

QY 139 GPEPHLAGDGPDKPSHLKLVFIHCRLRUPRISIDMESPLNL--SSEALSPATAKDE 196  
 Db 2206 --NSEISGQ-----MKPQLA--NMPISGRGMHIIHGVRNSSSTSVSKKGPP 2252

QY 197 ITQMLIKSAARSELGMMVVKSQEFLYARRRARRRKFAWKPVQIQTSEMKEVMEHTPMAYR 256  
 Db 2253 LKTPASKSSSEGQAT-TSPR-----GAKPSVK--SELSVAR-QTSQIGG 2294

QY 257 DSGSPPKNA--SPTSPGKPNISPROVSPVSPORSSPPQPAFPARTASKYSA 313  
 Db 2295 SSKASRSRSRSDTSPSRQPLSPRIOSPPGRNISGRRNGISPPNKLSQLRTSPSTA 2354

QY 257 DSGSPKNA--STPSLPGPKNTSPPROVSVRQSSPPKNSPPOPAFVARTASKYSA 313  
   | : | ; : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 2295 SSKAPRSRGSRSDTSRSPRQPLSRPQSPGRNSTSPGRNGISPPNKLSQLPRTSSPSTA 2354  
 QY 314 ASQVQRNRGNAKSLY 329  
   | : | ; : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 2355 ST---KSSGSKMSY 2366

RESULT 15  
 US-09-003-687A-7  
 Sequence 7, Application US/09003687A  
 Patent No. 5998600

GENERAL INFORMATION:

- APPLICANT: Barber, Nick
- APPLICANT: Clevers, Hans
- APPLICANT: Korinek, Vladimir
- APPLICANT: Morin, Patrice
- APPLICANT: Kinzler, Kenneth
- APPLICANT: Vogelstein, Bert
- APPLICANT: Sparks, Andrew

TITLE OF INVENTION: Beta Catenin, TCF-4, and APC

TITLE OF INVENTION: Interact to Prevent Cancer

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1001 G Street, N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ FOR Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/003,687A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/921,355

FILING DATE: 20-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A

REGISTRATION NUMBER: 32,145

REFERENCE/DOCKET NUMBER: 1107 . 05064

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

TELEX: 97430 BMB UT

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2973 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 5998600e

US-09-003-687A-7

Query Match 5.9%; Score 103; DB 2; Length 2973;  
 Best Local Similarity 22.3%; Pred. No. 0.52; Mismatches 107; Indels 46; Gaps 11;  
 Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;

QY 79 PFLEROPKVKHHRGSKILIPRFKSHGVHRELRSPHGHQSOSRVEDYHSTMQOLESTERNE 138  
 Db 2152 PQQEERFTSNKGFRILAKPGEKSTLTKEIESESKGIGKKV--YKSITGKVRS--- 2205

QY 139 GPEPHLAGDRPSKHLKLVFIRCLRLRIPRISDLMESPPIPNL--SGALSLSTATAKDE 196

Db 2206 -NSETSQ-----MOKPLQA-NMPSISRGRTWIMHIFGVRNNSSTSPVSKGP 2252

QY 197 ITOMILKSAARSELGMVSKRQEYLRRARRRKFAWKPVQLQSISEMKPVMVFHTPMAYR 256  
   | : | ; : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 2253 LKTPRASKSPSEGOTAT-TSPR-----GAKPSVK--SELSPVAR-QPSQIGG 2294

QY 257 DSGSPKNA--STPSLPGPKNTSPPROVSVRQSSPPKNSPPOPAFVARTASKYSA 313  
   | : | ; : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 2295 SSKAPRSRGSRSDTSRSPRQPLSRPQSPGRNSTSPGRNGISPPNKLSQLPRTSSPSTA 2354  
 QY 314 ASQVQRNRGNAKSLY 329  
   | : | ; : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 2355 ST---KSSGSKMSY 2366

Search completed: September 16, 2002, 00:43:15  
 Job time: 2724 sec

Mon Sep 16 08:49:20 2002

us-09-828-313-35.rai

Page 10

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

OM protein - protein search, using sw model

Run on: September 16, 2002, 00:34:21 ; Search time 42.86 seconds

(without alignments) 742.080 Million cell updates/sec

**title:** US-09-828-313-35  
**perfect score:** 1734  
**Sequence:** MGLTPFSCVTVQGYVRVWYPI.....SAASQOVRNRNGNAKSLYMA 331

**Scoring table:** BL05M62  
**Gapop** 10.0 , **Gapext** 0.5

**Searched:** 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database : **FOR PI.\*\***  
1: pix1:\*\*  
2: pix2:\*\*  
3: pix3:\*\*  
4: pix4:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

**RESULT** 1

T12533 hypothetical protein DKFZp434B034.1 - human (fragment)

C;Species: Homo sapiens (man)  
C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999  
C;Accession: T12533  
R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, June 1999  
A;Reference number: Z17524  
A;Accession: T12533  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Cross-references: EMBL: AL096728  
A;Experimental source: adult testis; clone DKFZp434B034.1  
A;Genetics:  
A;Note: DKFZp434B034.1

Query Match 7.6%; Score 131; DB 2; Length 413;  
Best Local Similarity 20.6%; Pred. No. 0.07%; Mismatches 142; Indels 102; Gaps 15;  
Matches 78; Conservative 57; Description

No.	Score	Query	Match	Length	DB	ID
1	131	7.6	413	2	T12533	hypothetical protein
2	124	7.2	795	2	E6654	hypothetical prote
3	118.5	6.8	1102	2	T26666	protein kinase C-C
4	117	6.7	2	T40503	protein kinase kin	
5	117	6.7	891	2	A38903	protein kinase 1 -
6	116	6.7	912	2	A54423	brevican precursor
7	114	6.6	742	2	F84643	hypothetical prote
8	112.5	6.5	576	2	T38293	hypothetical serin
9	111.5	6.4	539	2	T33658	hypothetical prote
10	111	6.4	202	2	T01510	hypothetical prote
11	109	6.3	651	2	T40505	hypothetical prote
12	108	6.2	1356	2	T11754	proline-rich prote
13	107	6.2	379	2	T05441	extensin-like prote
14	107	6.2	672	2	D85257	probable GT-like t
15	107	6.2	603	2	F98797	probable involve
16	107	6.2	1461	2	T41643	hypothetical prote
17	105.5	6.1	830	2	T18860	hypothetical prote
18	105	6.1	545	2	T36659	hypothetical prote
19	105	6.1	568	2	T36660	adenomatous polypo
20	105	6.0	2845	2	I45505	mucin 7 precursor,
21	104.5	6.0	377	2	A40018	serine/treonine k
22	104	6.0	597	2	AH2351	high molecular mas
23	104	6.0	1151	2	T18535	protein f6d22.14
24	104	6.0	2254	2	D84215	glucan 1,4-alpha-9
25	103.5	5.9	2022	1	T48818	alpha-fetoprotein
26	103	5.9	2783	1	A41948	adenomatous polyo
27	103	5.9	2843	1	R8HUP	probable extensin
28	102.5	5.9	2	C84005	hydroxyproline-ric	
29	102	5.9	620	2	S06733	

probable myosin I  
hypothetical prote  
unknown protein [1]  
hypothetical prote  
probable Inositol  
extensin - Volvox  
Csk1 protein FB28  
hypothetical prote  
hypothetical prote  
hypothetical prote  
unconventional myo

**RESULT 2**  
**E9654** hypothetical protein F16p17.12 [Imported] - Arabidopsis thaliana (mouse-ear cress)  
**C-Species:** Arabidopsis thaliana (mouse-ear cress)  
**C-Date:** 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
**C-Accesion:** E9654  
**R-Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; Ansen, N.F.; Hughes, B.; Huizar, L.**  
**Nature** 408: 815-820, 2000  
**A;Authors:** Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marzali, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Salano, H.  
**A;Author:** Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
**A;Title:** Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
**A;Reference number:** A86141; WJID:21016719  
**A;Accession:** E9654  
**A;Status:** preliminary  
**A;Molecule type:** DNA  
**A;Residues:** 1-966 <STO>  
**C;Genetics:**  
**A;Cross-references:** GB:AED005173; NID:98493585; PIDN:AAF75808.1; GSDB:GN00141  
**A;Gene:** FLP17.12  
**A;Map position:** 1

**Query Match 7.2%; Score 124; DB 2; Length 796; Best Local Similarity 23.8%; Pred. No: 0.14; Indels 52; Gaps 11; Matches 68; Conservative 36; Mismatches 132; Db 164 TPGSQSSQSPHHQQQSOPHHNNQNTWPSQFQNL-PFLHOPHQAVQTOQSQQRS QY 74 TPNAO-----PELEROPKVHRSKILPRESKHGKVHRELSRSPHSQSRK 120 Db 223 QFDQQLQQQQQASWRQVGOEHSGSGDWMKRPPVVEVRVLINLNATERRVSSRFAERE 222 QY 176 ESPLNOLSGEALSPATAKOBIT----OMILKSAARSEL---GMYVKRQFYLRARR 228 Db 283 RPPSFSKED----SSHRRPSELTMASKTPVPSVPHSLEVPYKPSVPPRQYQ----SSRS 337 QY 229 RKFAMWKPVQIQSSEMVKPMFHTPMAYRDG---SPPKMASTESLPGPKNISPROVSV 284 Db 338 SEAQOLSLLPSVSDSS----HASOPTRSNDOSHAVSKPQPVSKRPHPPMSQPPP--TSN 390 QY 285 PQRSSFPKRNVP-----PPQPAVARTASKYSAAQQVQRNR 322 Db 391 PFPPLSPSPNSKPFPMSSONSKRPPVPSOSOKSPKLVLVSQSQRSK 438**

**RESULT 4**  
**T40503** protein kinase kin1 - fission yeast (Schizosaccharomyces pombe)  
**C-Species:** Schizosaccharomyces pombe  
**C-Date:** 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Jan-2000  
**R;Gwilliam, R. Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.**  
**submitted to the EMBL Data Library, September 1998**  
**A;Reference number:** Z21933  
**A;Accession:** T40503  
**A;Status:** preliminary; translated from GB/EMBL/DBJ  
**A;Molecule type:** DNA  
**A;Residues:** 1-891 <GTM>  
**A;Cross-references:** EMBL:AL031154; PIDN:QAA20726.1; GSDB:GN00067; SPDB:SPBC4F6\_06  
**A;Experimental source:** strain 972h-; cosmid C4F6  
**C;Genetics:**  
**A;Map Position:** 2  
**C;Superfamily:** unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  

**Query Match 5.7%; Score 117; DB 2; Length 891; Best Local Similarity 21.8%; Pred. No: 0.53; Indels 92; Gaps 17; Matches 78; Conservative 50; Mismatches 137; Db 262 VHRDLKIENTLISKTDGKIDTFGLSNLYRRQSLRLRFGCSLLYAAPELLNAQPYTGE 321 QY 17 WVPDGHWEN--LSKSCSYHVLLGNDYY-----VCGSTPTVITNRMAE----- 60 Db 61 -EVLEVYCTVCATPNAQPFEROKVH---RGSKLIPFRSKHGVIVREARSPIHGS 115 QY 322 VDWWSGFLVVLVVGKRY-PFDQDNNSALHAKIKKGTVYESY-----LSDCKGL 371 QY 116 QOSRKVED-YVHSQMOLES---INN-ECPPEPLADRSKKHLKLVFTRHCLRALRUP 169 submitted to the EMBL Data Library, April 1997  
**A;Reference number:** Z20501  
**A;Accession:** T48666  
**A;Status:** preliminary; translated from GB/EMBL/DBJ  
**A;Molecule type:** mRNA  
**A;Residues:** 1-110 <MUB>  
**A;Cross-references:** EMBL:Y13104; NID:sl289874; PIDN:CAA73558.1  
**C;Superfamily:** unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog**

**Query Match 6.8%; Score 118.5; DB 2; Length 1102; Best Local Similarity 25.8%; Pred. No: 0.54; Indels 59; Gaps 18; Matches 85; Conservative 36; Mismatches 149; Db 473 FRIRAHANDDENPIPLSLSMMNTDIYAFHPLISIYLVSERRYBGGWNRAKPVSS 532**

RESULT	5	A; Molecule type: mRNA A; Residues: 1-912 <YAM> A; Cross-references: GB:X15887; NID:9452820; PIDN:CAA53481.1; PID:9452821 C; Superfamily: agrecan; C-type lectin homology; complement factor H repeat homology C; Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 24-Sep-1999 C; Accession: A38903; A36474 R; Levin, D.E.; Bishop, J.M. submitted to Genbank May 1991 A; Reference number: A38903 A; Accession: A38903 A; Molecule type: DNA A; Residues: 1-891 <LEV> A; Cross-references: GB:MG4999; NID:9173409; PIDN:AAA63577.1; PID:9173410 R; Levin, D.E.; Bishop, J.M. Proc. Natl. Acad. Sci. U.S.A. 87, 8272-8276, 1990 A; Title: A putative protein kinase gene (kin1(+)) is important for growth polarity in <i>Saccharomyces pombe</i> A; Reference number: A36474; MUID:9105979 A; Accession: A36474 A; Status: nucleic acid sequence not shown A; Molecule type: DNA A; Residues: 1-361 A; Cross-references: GB:M56060 C; Genetics: A; Gene: kin1+ C; Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog C; Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase F:123-395/Domain: protein kinase ATP-binding motif
Query Match	6.7%	Score 117; DB 2; Length 891;
Best Local Similarity	21.8%	Pred. No. 0.53;
Matches	78;	Conservative 50; Mismatches 137; Indels 92; Gaps 17;
QY	17	WVFDGHVEN--LSKCSVHDLIIGNPDYY-----VCGSPTPYTINRMAE-----60
Db	262	WHRDKEENILSKTGDKIILDFGLSNLYRQRSLRTRGSGLYFAAPELNAQIGE 321
QY	61	-EVLEYGVYTFVCAFPNAQFLERQPKVH---RGSKLIPRSKGVHVRERLPRTHOS 115
Db	322	VDWWSFGVLYVNLNGKV PFDQDNMAMAHAKIKKGTVWVPSI-----LSSDCKG 371
QY	116	QOSRKVFD-YHSVTMQLSES---IRN-EGEPBLAGDRPSKHLKVFTRHCLRALRLP 169
Db	372	LSRMLVYDPLKRATLEEVINHPMIRKVNVEGPASAPERSP-----412
QY	170	ISIDLMESPLPMLNSGALSPPTATAKDETTOMILKARS--ELGMY---VSKIQEFYL 222
Db	413	ITLPLDPPIIREMNGDFGPEKPTVRELKVISSEAYOSLAKTGFISGGPSADKKSFPE 472
QY	223	RRARRRKFANKPVGLOSIEMKPMHEHPTM-----AYRDSGSPKNASTP-S 269
Db	473	FRIKHAIDIENPILSLSANTDIDYARHPLISIVLVSERRVIEKGSGNWKARTPAS 532
QY	270	LPGPKNTSPPROPSVYORSSPPPKNVVPPPAFARTATSKYAAASQVORNGRNAK 326
Db	533	Vp----SSPVO-----PTSYNRILPPMPPEVV--AYKGDEESPRVSRNTSLAR 573
RESULT	6	best local precursor - bovine
Query Match	6.6%	Score 114; DB 2; Length 742;
Best Local Similarity	21.7%	Pred. No. 0.72;
Matches	60;	Conservative 48; Mismatches 127; Indels 42; Gaps 9;
QY	52	TITRNMAAEVLEYGVYTFVCAFPNAQFLERQPKVHVRERLSPKGVHVRERLSP 111
Db	352	TVFNOITANILQESLDGSPRSDRSLSLESALERVKERKTKLM-----ISENVVSP 404
QY	112	THGSPQSRTVFDVHSYMMQOLESTRNEGEPHLAGRSPRSHLKVFIRHCLRALRPLR 171
Db	405	DTSSPEK-----DMMSSHKSTAD--PNSILKYDESGRLRVNVIKSPRAV 456
QY	172	IDLMESPPLNLSGEALSPPTATAKDETTOMILKARS--VYSRQEYLRA 225
Db	457	OSPVTSPLNRSPTGSPASISRFH-----SSPSLGLTSILHDHGSKDEESTSS 508
QY	226	RRRKFANKPVGLOSIEMKPMHEHPTM-----HTPMAYRDGSPPKNASTPSLPPKNTSPPR 281
A; Status: preliminary		

**RESULT 8**

Query Match 6.5%; Score 112.5; DB 2; Length 576;  
Best Local Similarity 23.8%; Pred No. 0; Indels 68; Gaps 14;  
Matches 75; Conservative 35; Mismatches 120; Indexes 85; Gaps 14;

OY 67 VTYFVCPATPNAQPFLERQPKVVKHGSKILPRFSKGKV-HYRELRSPTGSDQSRSKVDYH 125  
Db 18 ITGKICFEINA-----ATSPRSFQQLQRIGYAVHGFLLSPSKHKPTIASDSD 69

OY 126 SVTMMQLESIRNEGPEPHLAGDRPSKHL-----KAVFTRHCLR 163  
Db 70 EAT---EFTHSQPLPPIAGPSSKSHADFKFEPSSKASSLPCSKNSDNICIMLK 125

OY 164 ALRLPRIST-----DAMESPLPNISGEALSLPT----ATAK-----DEITOMILK 203

Db 126 ATISKRYARLGSSQSAKAKLMVP-PNLGKPELUSANSFSFSSAKTFOLPEFDSTV---P 180

OY 204 SAARSELGMVSKRQEYLRHARRRKPAWKPVLOISEKPKVMEFHPPM-----AYR 256  
Db 181 SASLYKQDFNSNPAPITTSATHTSQF-----STSSSSVNSVHPVPMVNPYFOYN 233

OY 257 DSGSIPKIASTPSPGPKNITSPPRQSYVQRSSP---PKKVSPPQQPAFVARTASYSA 314

Db 234 SSMTRAPSSESSSSVAP---EVPRROFSVSSASDPPQTPISSMSPPIPP---TPSOFSAF 284

OY 315 SQQVNORNRAKSLY 329  
Db 285 MYQNQQSUSPQSHY 299

**RESULT 9**

Query Match 6.4%; Score 111; DB 2; Length 539;  
Best Local Similarity 23.0%; Pred No. 0; Indels 74; Gaps 6;  
Matches 43; Conservative 38; Mismatches 75; Indexes 31; Gaps 6;

OY 154 KLVFTRHCLRALRPLRISID-LMESPLPNLSGALSPTATAKDETQMTLKSARSELGM 212  
Db 300 KRVYNORRERFAERRRMEEEMILAMPICRNGEAL-----DRAQIVYRNKAEAR 352

OY 213 YVSKRQEFYLLRARRRKPAWKPVQLOISEMVKPMETHTPMAYRUDSGSPKNASTPSLG 272

Db 353 LRYQRMTPDQPKSYNQKRYTKRKRNEADMISAGO---MAVSSGCTTMGTGTTAKKD 408

OY 273 P-----KNTSPPROVSVIPOR--SSPPKVNSSPPQPAFVARTASYSA 317

Db 409 DMDAQLSSLERDVKRTQHADQILLRQANQSCPTPSRVTPPPQVVIINGQOHH---QQ 464

OY 318 VORNEN 324  
Db 465 LVGNQGN 471

**RESULT 10**

Query Match 6.4%; Score 111.5; DB 2; Length 539;  
Best Local Similarity 23.0%; Pred No. 0; Indels 74; Gaps 6;  
Matches 43; Conservative 38; Mismatches 75; Indexes 31; Gaps 6;

OY T01510 hypothetical protein T10M13.10 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse/ear cross)  
C;Date: 19-Feb-1999 #sequence\_revision 19-Feb-1999 #text\_change 04-Mar-2000  
C;accession: T01510  
R;Johnson, A. F.; de la Bastide, M.; Lohdi, M.; Hoffman, J. J.; Hasegawa, A.; Gnoj, I.; G  
Martienssen, R.; McCombie, W.  
A;Description: The sequence of the Arabidopsis thaliana T10M13 BAC.  
submitted to the EMBL Data Library, May 1997  
A;reference number: Z11346  
A;status: submitted from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-202 <OH>  
A;Cross-references: EMBL:AF001308; NID:92104523; PID:92104533  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Map position: 45  
A;Note: T10M13.10  
C;Superfamily: Arabidopsis thaliana hypothetical protein T10M13.10

Query Match 6.4%; Score 111; DB 2; Length 202;  
Best Local Similarity 21.9%; Pred No. 0; Indels 23; Gaps 7;  
Matches 54; Conservative 24; Mismatches 89; Indexes 80; Gaps 7;

OY 16 RVVFDGHENLSKSCVHDLLGDPDVCCSTP-----YTINRMA---AEEVL 63  
Db 19 KVMSDGRVYQNLSEETTVAEMLLENPOHIVVFPDSISFNNDAKTVKRKLAPLPADKT 78

OY 64 EIGVTFVCPATPNAQPFLERQPKVVKHGSKILPRFSKGKV-HYRELRSPTGSDQSRSKVDYH 123  
Db 79 ERGKTYL-----PSATAMRSFSFYEGILWFT 135

OY 184 GEALSPTATAKDITOMILKSAAMSSELGMVSKRQEYLRHARRRKPAWKVULSISEM 243

Db 136 TRSYKNNPATD---TVAATSVGRLEAEMEEDRPEFLSQLSGR---GKWPSDLPIREK 190

OY 244 KPVMFH 250  
Db 191 KAKKIH 197

**RESULT 9**

Query Match 6.4%; Score 111; DB 2; Length 539;  
Best Local Similarity 23.0%; Pred No. 0; Indels 74; Gaps 6;  
Matches 43; Conservative 38; Mismatches 75; Indexes 31; Gaps 6;

OY T01510 hypothetical protein T10M13.10 - Arabidopsis thaliana  
C;Species: Caenorhabditis elegans  
C;Accession: T33658  
R;Wohlbomann, P.; Bauer, C.; Rohlfing, T.; Gilliam, B.  
A;Description: The sequence of C. elegans cosmid W07B3.  
A;Reference number: Z21384  
A;Accession: T33658  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-539 <WHO>  
A;Cross-references: EMBL:AF100304; PIDN: AAC68911.1; GSPDB: GN00021; CBSP: W07B3.2c  
A;Experimental source: strain Bristol N2; clone W07B3

RESULT 11	T16K5.190 - Arabidopsis thaliana	hypothetical protein T16K5.190 - Arabidopsis thaliana (mouse-ear cress)	C;Species: Arabidopsis thaliana (mouse-ear cress)	C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000	C;Accession: T465050	R;Rieder, M.; Gabel, C.; Mueller-Auer, S.; Schaeffer, M.; Zipp, M.; Meewes, H.W.; Lemcke, submitted to the Protein Sequence Database, January 2000	A;Accession: T465050	A;Status: preliminary	A;Molecule type: DNA	A;Residues: 1-651 <RIE>	A;Cross-references: EMBL:AL122965	A;Experimental source: cultivar Columbia; BAC clone T16K5	C;Genetics:	A;Map position: 3	A;Introns: 52/1; 83/1; 153/2; 209/1; 257/3; 278/3; 309/3; 363/3; 424/1; 479/1; 501/1; 54 A;Note: T16K5.190	
Query Match	6.3%	Score 109; DB 2; Length 651;	Best Local Similarity 22.5%; Pred. No. 1.5;	Matches 62; Conservative 30; Mismatches 92; Indels 92; Gaps 13;	Qy 52 TITPMRAEEVIEVYGVYFVCATPNAQPFLERQPKVWRGS-KILPRSKHVGVRHREL 109	Db 387 SISRKMLLAQRIVEQDDETSCS-----POLIATSTASPRLPRSSNNKKRTVPL 435	Qy 110 -----SPPHGQSOSKRKVDFYHSVTMQLSERINRNGPPELRPSKHLKFIRHCLR 163	Db 436 ESDIFPPSFHKPEER-----KTKQELQFKTKESTE-----N 469	Qy 164 ALRERFISIDLMESPPLNPISGEALSPTATADETOMILKSAR--SEGMVYWSKQEF 220	Db 470 ALK-----ELVSSPEGSGSE-----KQARGTSEPVGECTKKIKO- 506	Qy 221 YLRRARRRKKFWKPKVLQLSIEMKPKWEEHPHM--AYRDSGSPPKNASTPS-LPGPKN 275	Db 507 -----RVYXKERNPKPDKSTIEKKKSYQDQHRSVAPPQGYPKEGGPPQGYPPAG 561	Qy 276 1SPPR-QVSVPQRSSPPK-----WVSPQQ 300	Db 562 YPPPOQYPOAGYPAGYPPOQGYQGQGYPAQGYPQQ 597	Qy 305 ARTASKYSASQOVNRNRGAKS 327	Db 652 ---EEWEFDNBISGGGSAQT 670
RESULT 13	T0541	proline-rich protein F7K2.50 - Arabidopsis thaliana (fragment)	C;Species: Arabidopsis thaliana (mouse-ear cress)	C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999	C;Accession: T05441	R;Bovari, M.; Wedderburn, H.; Wambutt, R.; Barcroft, I.; Meewes, H.W.; Mayer, K.F.X.; Schue submitted to the Protein Sequence Database, November 1998	A;Reference number: 215416	A;Accession: T05441	A;Molecule type: DNA	A;Residues: 1-379 <BEV>	A;Cross-references: EMBL:AL033545	A;Experimental source: cultivar Columbia; BAC clone F7K2	C;Genetics:	A;Map position: 4	A;Note: F7K2.50	
Query Match	6.3%	Score 107; DB 2; Length 379;	Best Local Similarity 48.9%; Pred. No. 1;	Matches 22; Conservative 6; Mismatches 15; Indels 2; Gaps 1;	Qy 258 SGSPPKNASTPSLPGPKNISPPROVSYVQRSSPPKNSPPOPA 302	Db 79 SNSPPVYATTPALPKPLPP--LSPQQTTPPPPA 121	Query Match	6.3%	Score 107; DB 2; Length 379;	Best Local Similarity 48.9%; Pred. No. 1;	Matches 22; Conservative 6; Mismatches 15; Indels 2; Gaps 1;	Qy 258 SGSPPKNASTPSLPGPKNISPPROVSYVQRSSPPKNSPPOPA 302	Db 79 SNSPPVYATTPALPKPLPP--LSPQQTTPPPPA 121			
RESULT 14	D85257	extensin-like protein [imported] - Arabidopsis thaliana	C;Species: Arabidopsis thaliana (mouse-ear cress)	C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001	C;Accession: D85257	R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold SP Nature 402, 769-777, 1999	A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.	A;Reference number: A85001; MUID:20083488	A;Accession: D85257	A;Status: preliminary	A;Molecule type: DNA	A;Residues: 1-379 <STO>	A;Cross-references: GB:NC_001268; NID:97269093; PIDN:CAB79202.1; GSDB:GN00140	C;Genetics:	A;Gene: Atvq22470	A;Map position: 4
Query Match	6.2%	Score 108; DB 2; Length 1356;	Best Local Similarity 21.7%; Pred. No. 4.4;	Matches 70; Conservative 46; Mismatches 101; Indels 106; Gaps 17;	Qy 73 ATPNAQPEL-----FROPKVHRGSKILPPRS-----KHCVHYRELRSPTHS-O 115	Db 386 ANPHAPKTTAAQKERRQERGP-----LPKWDIEAGDDNDYWMILMEAQYSQYGAQ 436	Qy 117 QSKYFDFH-----SVMQOLESRINRNEGPEPHLGDREPKSHKLKFIRHCLRALRPRISIDL 174	Db 437 OFEHOLDHQASILQQAQ-----HMASFAPFR-----PMFLSPLH 475	Qy 175 MESPL-----PNISGEALS--PTA-----TAKEDETONILKSAARSEAGMYSK- 216	Db 476 MSPLDROGGVTPSVNGEINTAIPAPIVNPPTAPRPVTDITLSEYVKQLEYVSEE 535	Qy 217 --RQFYLKARRRKFPAWKVQVLSIBMKVYMERHTPM--AYRDSG---SP----- 261	Db 536 NLQKOFFURKMGPECYLVIALASFRYRLLDYLSEALKDSTKVMSPDGLORA 595	Qy 262 -----PKNASTPSLGPKNISP-----PROVSVPDRSSPPKAVSPPPOPQAFV 304	Db 596 PVNPPTIWPLMVGADSLGPSSQPOQFQRONGPAATRAPVESQOASS-SKRPQDQ---- 651		

Query	Match	Score	Length
Best	Local	Similarity	DB
Matches	22;	Conservative	2;
		48.9%	
		6;	Mismatches
			15;
			Indels
			2;
			Gaps
258	S S P K N A S T P S L G P K N I S P P R Q V S V P O R S S P P K N V S P P Q Q A	302	
79	S T S P P V A T P R P A T P R P P K L P P - : - S P P Q V P P P P P A T P R P P A	121	

RESULT  
F96797

Search completed: September 16, 2002, 00:44:16  
Job time: 595 sec

Scoring table:		BLOSUM62				ALIGMENTS				
Searched:	105224 seqs., 38719550 residues	Gapop	10.0	Gapext.	0.5					
Total number of hits satisfying chosen parameters:	105224									
Minimum DB seq. length:	0									
Maximum DB seq. length:	200000000									
Post-processing:	Minimum Match 0%									
	Maximum Match 100%									
	Listing first 45 summaries									
Database :	Swissprot; 40.4									
<b>Pred.</b> No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.										
Result No.	Score	Match	Length	DB	ID	Description				
1	117	6.7	891	1	KINL_SCHPO	RESULT 1				
2	116	6.7	912	1	PGCB_BOVIN	KINL_SCHPO STANDARD;	PRT;	891 AA.		
3	112.5	6.5	569	1	STET_SCHPO	P22987 schizosaccharomyces pombe kinase				
4	105	6.1	2845	1	APC_MOUSE	P22987; 074392; 01-AUG-1991 (Rel. 19, Created)				
5	103.5	6.0	830	1	DYNL_CAEEL	DT 16-OCT-2001 (Rel. 40, Last sequence update)				
6	103	5.9	1822	1	ZAP3_HUMAN	DT 16-OCT-2001 (Rel. 40, Last annotation update)				
7	103	5.9	2843	1	APC_HUMAN	DE Protein kinase kin1 (EC 2.7.1.-).				
8	103	5.9	3703	1	ABFL1_HUMAN	OS KIN1 OR SP64F6.06.				
9	102.5	5.9	864	1	WS14_MOUSE	ON Schizosaccharomyces pombe (Fission yeast).				
10	102	5.9	620	1	EXTN_TOBAC	OC Fungi: Ascomycota; Schizosaccharomycetes; Schizosaccharomyces pombe.				
11	101.5	5.9	1321	1	IRS2_MOUSE	OC Schizosaccharomyces pombe; Schizosaccharomyces pombe.				
12	101	5.8	1943	1	RWI_DROME	OC NCBI_TaxID=4896;				
13	100	5.8	1943	1	PC15_MOUSE	RP SEQUENCE FROM N.A.				
14	100	5.8	2842	1	APC_RAT	RN MEDLINE=91045979; PubMed=2236039;				
15	100	5.8	3726	1	APC1_MOUSE	RA Lexin D.E., Bishop J.M.;				
16	99.5	5.7	760	1	EXMT1_MOUSE	RT "A putative protein kinase gene (kin1+) is important for growth polarity in Schizosaccharomyces pombe";				
17	99	5.7	555	1	GPM1_CHLRE	RL Proc. Natl Acad. Sci. U.S.A. 87:8272-8276(1990).				
18	99	5.7	759	1	F2ML1_RAT	RN [2]				
19	99	5.7	1021	1	YL18_CAEEL	RP STRAIN=972;				
20	99	5.7	3530	1	MY15_HUMAN	RA William R. Rajandream M.N., Barrell B.G., Shelton J., Churcher C.M.;				
21	98.5	5.7	806	1	MK077_MOUSE	RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.				
22	98	5.7	497	1	WAS2_HUMAN	CC -1- FUNCTION: PROBABLE SERINE/THREONINE PROTEIN KINASE. IMPORTANT FOR GROWTH POLARITY IN S. POMBE.				
23	98	5.7	678	1	T2D5_MOUSE	CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.				
24	98	5.7	678	1	T2D5_RAT	CC -1- STRONG, TO YEAST KIN1 AND KIN2.				
25	97.5	5.6	1745	1	Z011_MOUSE	CC -----				
26	97	5.6	306	1	EXTN_DAUCA	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_licensing@ib-sib.ch.html">http://www.isb-sib.ch/announce/or_send_an_email_to_licensing@ib-sib.ch.html</a> ).				
27	97	5.6	426	1	EXLP_TOBAC	CC -----				
28	97	5.6	630	1	Z147_HUMAN	CC -----				
29	97	5.6	677	1	T2D5_HUMAN	DR EMBL: M61999; AAA3577.1; -.				
30	97	5.6	1157	1	Y182_HUMAN	DR P97691; AL031534; CAA20726.1; -.				
31	96.5	5.6	1386	1	ZAP3_MOUSE	DR P97692; A38903.				
32	96	5.5	1487	1	BLM_DROME	DR HSP; P00518; IPKB.				
33	96	5.5	1805	1	RWL_HUMAN	DR InterPro: IPRO00719; Euk_Pkinase.				
						DR InterPro: IPRO01722; Kal.				
						DR InterPro: IPRO02290; Ser_Thr_Pkinase.				
						DR InterPro: IPRO01245; Tyr_Pkinase.				
						DR Pfam: PF00249; Kal_1.				
						DR Pfam: PF00218; IPKB.				
						DR HSP; P00518; IPKB.				
						DR Prints; PR00109; TTKINASE.				
						DR Smart; SW0220; S_TKC; I.				
						DR Prosite; PS00107; PROTEIN_KINASE_ATP; I.				
						DR Prosite; PS00108; PROTEIN_KINASE_ST; I.				
						DR Prosite; PS50011; PROTEIN_KINASE_DOM; I.				
						DR KW Transferase; Serine/threonine-protein kinase; ATP-binding.				
						FT DOMAIN; PROTEIN_KINASE.				
						FT DOMAIN; PROTEIN_KINASE.				
						FT 125 395 PROTEIN_KINASE.				



Qy 184 GEALSPATAKDEITQMLKSAARSELGMVSKRQEFLRRARRRK-----FA 232  
 | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 425 AEA--PRHLLEFE-TOSTWPPGLGSSEBEGVLEEEKEEEREVEEDEALMA 481  
 | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Qy 233 WKPVLISIEMKPMVFHPTMAYRDGGSPKNA-----STPSLPGPKNSPPROVSVPQ 286  
 | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 482 WISELSLSDPBPAPLPTEPVPEBSLTQASPVYRALQPSVSPPYDPDEAPRPPRVLGPPT 541  
 | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Qy 287 RSSPPKKN---VSPPQPAFVARTASKYSASOOVQRBRGNAK 326  
 :: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 542 KTLPLTPREGNLASPPPLVLAGARETEETGPGELSGAPRGSE 584  
 | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 RESULT 3  
 STE7\_SCHPO STANDARD; PRT; 569 AA.  
 ID STE7\_SCHPO STANDARD; PRT; 569 AA.  
 AC Q10135;  
 DT 01-FEB-1995 (Rel. 33, Created)  
 DT 01-FEB-1995 (Rel. 33, Last sequence update)  
 DE Ste7 protein.  
 GN STE7 OR SPC23E2\_03C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=2096734; PubMed=10833379;  
 RA Matsuyama A., Yabana N., Watarabe Y., Yamamoto M.;  
 RT "Schizosaccharomyces pombe Ste7p is required for both promotion and  
 RL withholding of the entry to meiosis.";  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;  
 RL CC Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: HAS A ROLE IN PROMOTING MEIOSIS WHEREBY IT  
 CC IS INVOLVED IN ESTABLISHING THE MATING PHEROMONE SIGNALING  
 CC PATHWAY. IT ALSO HAS A ROLE IN SUPPRESSING MEIOSIS UNTIL THE  
 CC CONJUGATION PROCESS IS COMPLETE.  
 CC -!- INDUCTION: BY NITROGEN STARVATION.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 DR EMBL; AB036789; BAR90541.1; -.  
 DR EMBL; Z6887; CAA3115.3; -.  
 KW Conjugation; Meiosis; Pheromone.  
 FT DOMAIN 261 SER-RICH.  
 FT DOMAIN 464 535 SER-RICH.  
 SQ SEQUENCE 569 AA; 61101 MW; FDD23893E9A7A6AC CRC64;

Query Match Score 6.5%; Score 112.5; DB 1; Length 569;  
 Best Local Similarity 23.8%; Pred. No. 0.37; Mismatches 75; Conservative 35; Mismatches 120; Indels 85; Gaps 14;

Qy 67 VTFVFCATPNQDFPLERQPKVVRHRSKILPRESKHGV-HYRELRSPTHGQSQRKVFDYH 125  
 | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 18 ITGKTCFENNA-----ASPSFSQFOILTRIGKAVYHNGFLSPSMKHPTASDS 69  
 | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Qy 126 SVTMQQLSIRNNEGPERPHAGDRSKL-----KLVFIRHCLR 163  
 | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |  
 Db 70 EAT---EIFTHSQPLIPAGPSSKSHADFKFKFKSKASSLPCSKNSDMSAVNCYMLK 125  
 | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |

RESULT 4  
 APC\_MOUSE STANDARD; PRT; 2845 AA.  
 ID APC\_MOUSE STANDARD; PRT; 2845 AA.  
 AC Q61315; 060404;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Adenomatous polyposis coli protein (APC protein) (mAPC).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBL-TaxID=1090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.  
 RC STRAIN=C57BL/6J; TISSUE=Brain;  
 RX MEDLINE=92963101; PubMed=1351018;  
 RA SU L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
 RA Luongo C., Goud K.A., Dove W.F.;  
 RT Multiple intestinal neoplasia caused by a mutation in the murine  
 RT homolog of the APC gene;  
 RL Science 256:668-670(1992).  
 RN [2]  
 RP ERATUM.  
 RA SU L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
 RA Luongo C., Gould K.A., Dove W.F.;  
 RL Science 256:1114-1114(1992).  
 RN [3]  
 RP SEQUENCE OF 1-15 FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=Liver;  
 RA Dicker F., Lambert S., Reitmair A., Ballhausen W.G.;  
 RT "The murine APC gene: alternative splicing of 5' untranslated  
 RT region segments";  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE=94061824; PubMed=8242607;  
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;  
 RT "APC gene messenger RNA: novel isoforms that lack exon 7.;"  
 RL Cancer Res. 53:5589-5591(1993).  
 CC -!- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY).

CC -!- SIMILARITY: FORMS HOMOOLIGOMERS AND ASSOCIATES WITH CATENINS (BY SIMILARITY).

CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE PRODUCED BY ALTERNATIVE SPlicing.

CC -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLINE, KIDNEY, HEART, LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.

CC -!- PTM: PHOSPHORYLATION BY GSK-3B (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 7 ARM REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

---

CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).

---

DR InterPro: IPR000225; Armadillo.

DR SMART, SW00185; ARM\_5.

DR PROSITE; PS00166; ARM\_REPEAT; 1.

KW Anti-oncogene; phosphorylation; Alternative splicing; Repeat; Coiled coil.

FT DOMAIN 1 61 COILED COIL (POTENTIAL).

FT DOMAIN 125 245 COILED COIL (POTENTIAL).

FT DOMAIN 1 728 LEU-RICH.

FT REPEAT 451 493 ARM 1.

FT REPEAT 503 545 ARM 2.

FT REPEAT 546 589 ARM 3.

FT REPEAT 590 636 ARM 4.

FT REPEAT 637 681 ARM 5.

FT REPEAT 682 723 ARM 6.

FT REPEAT 724 765 ARM 7.

FT DOMAIN 739 2834 SER-RICH.

FT DOMAIN 1130 1156 ASP(GLU)-RICH (ACIDIC).

FT DOMAIN 1556 1575 ASP(GLU)-RICH (ACIDIC).

FT DOMAIN 1864 1891 HIGHLY CHARGED.

FT VARPLIC 243 276 MISSING (IN ISOFORM 2 AND ISOFORM 4).

FT VARPLIC 310 410 MISSING (IN ISOFORM 3 AND ISOFORM 4).

FT VARIANT 120 120 T -> A (IN STRAIN CASTEII).

FT VARIANT 493 493 V -> I (IN STRAIN CASTEII).

FT VARIANT 797 797 Y -> F (IN STRAIN CASTEII).

FT VARIANT 1330 1330 A -> T (IN STRAIN CASTEII).

FT VARIANT 1618 1618 S (IN STRAIN CASTEII).

FT VARIANT 2294 2294 G -> A (IN STRAIN CASTEII).

FT VARIANT 2496 2496 H -> Q (IN STRAIN CASTEII).

FT VARIANT 2523 2523 T -> A (IN STRAIN CASTEII).

FT VARIANT 2813 2813 T -> S (IN STRAIN CASTEII).

SQ SEQUENCE 2845 AA; 311086 MW; 145CAT3CF570A99 CRC64;

Query Match 6.1%; Score 105; DB 1; Length 2845;

Best Local Similarity 21.7%; Pred. No. 10; Indels 50; Gaps 9; Matches 56; Conservative 42; Mismatches 110; Delins 50; Gaps 9;

Qy | 79 PFLURROPPVWRGSKTLPPFKHGKVHRELSPTHGSOOSRKVFDYHSVTMQLSES--I 135

Db 2152 PDQEKEPKTSNKGKPRLKGEKSTLEAKKIRESENKGKVK-YKSLTGKIKNSRI 2209

Qy | 136 RNEEPEPPILAGDRPSKHLVLFTRHCLRALUPRISDEMESPLNU--SCEALSPATA 193

Db 2210 SSQKQP-----LPTNNPMSISRGRTMHRPHGIRNNSSSTPVSK 2249

Qy | 194 KDEITOMILKSAARSELGLMVVKROEFYLRARRRKFAWKPVQLQSISEMPVPMFHTM 253

Db 2250 GPPLKTPASKPSSEG-GATTSPR-----GTPKA--GKSELSPLTQTSQI 2292

Qy | 254 AYRQSGSPKNA--SPPSLPGPKNSIPSPRQYSPQRSSPPPKVNSPPRQPFVARTASY 311

Db 2293 SGSKNGKSSRSGRSDSIPSPRTOQPLSRPMQSPGRNSISPGRNGISPNNKLSQLPRTSSPS 2352

Qy | 312 SAASQVQVRNRNAKLY 329

Db 2353 TAST-----KSSGSGKMSY 2366

DR PROSITE; PS00140; DYNAMIN\_1.

DR PROSITE; PS50003; PH\_DOMAIN\_1.

DR PROSITE; PS00169; PH; 1.

DR PRINS; PRO0195; DYNAMIN.

DR SMART; SM00053; DYNC; 1.

DR SMART; SM00302; GED; 1.

DR SMART; SM00233; PH; 1.

DR PROSITE; PS00410; DYNAMIN\_1.

DR PROSITE; PS50003; PH\_DOMAIN\_1.

KW Hydrolase; Motor protein; GTP-binding; Microtubules; Multigene family; Endocytosis.

FT NP\_BIND 40 47 GTP (BY SIMILARITY).

FT NP\_BIND 138 142 GTP (BY SIMILARITY).

FT NP\_BIND 207 210 GTP (BY SIMILARITY).

FT DOMAIN 519 624 PH.

SQ SEQUENCE 830 AA; 93348 MW; FC3D7106D079EDC5 CRC64;

Query Match 6.0%; Score 103.5; DB 1; Length 830;

Best Local Similarity 24.2%; Pred. No. 2.8; Indels 47; Gaps 10; Matches 52; Conservative 32; Mismatches 84; Delins 47; Gaps 10;

Qy | 98 FSKHGKVHRELSPTHGSOOSRKVFDYHSVTMQLSESIRNEGEPHLAGDRPSKHLV 157

Db 619 FLRAGVYPEKOKAQEDQEMEDTSIDQBLERQVETIRN-----LVDSY 663

Qy | 158 IRHCIRALR-LPRISIDLMHESPIPNLSEGAALSPTATAKEDITOMILK---SAARSELG 211

DT 30-MAY-2000 (Rel. 39, last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)

DR Dynamin (EC 3.6.1.50).

GN DYN-1.

OS Caenorhabditis elegans.

OC Bukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peioderinae; caenorhabditis.

NCBL\_TAXID=6239.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=97439883; PUBMED=994229;

RA Clark S.G., Shurland D.L., Meyerowitz E.M., Bargmann C.I., van der Bliek A.M.;

RT "A dynamin GTPase mutation causes a rapid and reversible temperature-induced locomotion defect in *C. elegans*." (1997).

RN [2]

RP REVISONS TO C-TERMINUS.

RC STRAIN=BRISTOL N2;

RA van der Bliek A.M.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

CC -|- Function: microtubule-associated force-producing protein involved in producing microtubule bundles and able to bind and hydrolyze GTP. Most probably involved in vesicular trafficking processes, in particular endocytosis.

CC -|- Catalytic activity: GTP + H<sub>2</sub>O = GDP + phosphate.

CC -|- Subcellular location: microtubule-associated.

CC -|- Similarity: belongs to the dynamin family.

CC -|- Similarity: contains 1 PH domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR HSSP; L29031; ABT2228.2.;

DR InterPro; IPR00140; Dynamin.

DR InterPro; IPR00075; Dynamin\_Central.

DR InterPro; IPR00130; GED.

DR InterPro; IPR001849; PH.

DR Pfam; PF00350; dynamin\_1.

DR Pfam; PF01031; dynamin\_2.

DR Pfam; PF02212; GED.

DR Pfam; PF0169; PH; 1.

DR PRINS; PRO0195; DYNAMIN.

DR SMART; SM00053; DYNC; 1.

DR SMART; SM00302; GED; 1.

DR SMART; SM00233; PH; 1.

DR PROSITE; PS00410; DYNAMIN\_1.

DR PROSITE; PS50003; PH\_DOMAIN\_1.



- RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients";  
RT Science 253:655-659(1991).  
RN [6]
- RP VARIANT FAP  
RX MEDLINE=93265030; PubMed=1338904;  
RA M1yoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T., Miki T., Mori T., Nakamura Y.;  
RA "Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene.";  
RL Hum. Mol. Genet. 1:229-233(1992).  
RN [7]
- RP VARIANT FAP  
RX MEDLINE=93241793; PubMed=1338691;  
RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y., Nakamura Y., Horii A.;  
RA "Somatic mutation of the APC gene in gastric cancer: frequent mutations in very well differentiated adenocarcinoma and signet-ring cell carcinoma.";  
RL Hum. Mol. Genet. 1:559-563(1992).  
RN [8]
- RP VARIANT FAP W-1348, AND VARIANTS D-1118; M-1292; V-1304 AND S-2502.  
RX MEDLINE=93250848; PubMed=1338764;  
RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M., Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J., Baba S., Nakamura Y.;  
RA "Screening for germ-line mutations in familial adenomatous polyposis patients: 61 new patients and a summary of 150 unrelated patients.";  
RL Hum. Mutat. 1:467-473(1992).  
RN [9]
- RP VARIANT FAP TRP-99.  
RC TISSUE=Peripheral blood lymphocytes;  
RX MEDLINE=9513544; PubMed=783149;  
RA Dobbie Z., Spycher M., Huerlimann R., Ammann R., Ammann T., Roth J., Mueller S., Mueller H., Scott R.J.;  
RA "Mutational analysis of the first 14 exons of the adenomatous polyposis coli (APC) gene";  
RL Eur. J. Cancer 30A:1709-1713(1994).  
RN [10]
- RP VARIANT FAP GLY-172.  
RX MEDLINE=95135430; PubMed=7833931;  
RA Stella A., Montero M., Resta N., Marchese C., Suca F., Gentile M., Romio L., Pilia S., Prete F., Mareni C., Guanti G.;  
RA "Four novel mutations of the APC (adenomatous polyposis coli) gene in FAP patients";  
RL Hum. Mol. Genet. 3:1687-1688(1994).  
RN [11]
- RP BERRATUM.  
RA Stella A., Montero M., Resta N., Marchese C., Suca F., Gentile M., Romio L., Pilia S., Prete F., Mareni C., Guanti G.;  
RL Hum. Mol. Genet. 3:1918-1918(1994).  
RN [12]
- RP VARIANT FAP TLE-171.  
RX MEDLINE=97144176; PubMed=899002;  
RA van der Luijt R.B., Meera Khan P., Vassena H.F.A., Tops C.M.J., van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M., Plug R.J., Griffioen G., Fodde R.;  
RA "Molecular analysis of the APC gene in 105 Dutch kindreds with familial adenomatous polyposis: 67 germline mutations identified by RT DGGE, PTT, and southern analysis";  
RL Hum. Mutat. 9:7-16(1997).  
RN [13]
- RP VARIANT LYS-1307.  
RX MEDLINE=98040548; PubMed=9731522;  
RA Redston M., Mathewson K.L., Yuan Z.Q., Neuhansen S.L., Satagopan J., Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H., Anton-Ozeckis D., Andriulis I., Daly M., Pinsky L., Schrag D., Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C., Gadwin A., Warner E., Weber W., Foukakis K.;  
RA "The APC 11307K allele and breast cancer risk";  
RL Natl. Genet. 20:13-14(1998).  
RN [14]
- RP VARIANTS LYS-1307 AND GLN-1317.
- RC TISSUE=Peripheral blood;  
RX MEDLINE=98393712; PubMed=9724771;  
RA Fraying I.M., Beck N.E., Illyas M., Dove-Edwin I., Goodman P., Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W., Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;  
RA "The APC variants 11307K and E131Q are associated with colorectal tumors, but not always with a family history";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).  
RN [15]
- RP VARIANT LYS-1307.  
RX MEDLINE=98400259; PubMed=9731533;  
RA Woodge T., King S.M., Wacholder S., Hartge P., Struwing J.P., McDonald M., Laden S.J., Tucker M.A., Brody L.C.;  
RA "Inherited colorectal polyposis and cancer risk of the APC 11307K polymorphism";  
RT Ashkenazi Jevs.;  
RL Ann. J. Hum. Genet. 64:378-384(1999).  
RN [16]
- RP VARIANT LYS-1307.  
RX MEDLINE=99138651; PubMed=9727276;  
RA Gryfe R., Di Nicola N., Lal G., Gallinger S., Redston M.;  
RA "Inherited colorectal polyposis and cancer risk of the APC 11307K polymorphism";  
RT Wallis Y.L., Morton D.G., McTeigue C.M., Macdonald F.;  
RT "Molecular analysis of the APC gene in 205 families: extended genotype-phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition.;"  
RT APC amino acid changes in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition.;"  
RT APC amino acid changes in colorectal cancer predisposition.;"  
RN [17]
- RP VARIANT FAP CYS-1171 AND THR-2738, AND VARIANTS GLY-1057 AND VAL-1822.  
RX MEDLINE=99133859; PubMed=9590360;  
RA Wallis Y.L., Morton D.G., McTeigue C.M., Macdonald F.;  
RT "Molecular analysis of the APC gene in 205 families: extended genotype-phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition.;"  
RT APC amino acid changes in colorectal cancer predisposition.;"  
RN [18]
- RP VARIANT FAP PRO-184.  
RX MEDLINE=99401091; PubMed=10470088;  
RA Lamblin H., Illyas M., Rowan A., Clark S., Johnson V., Bell J.A., Frayling I.M., Eftathliou J., Pack K., Payne S., Roylance R., Gorman P., Sheet D., Neale K., Phillips R., Talbot I.C., Bodmer W.F., Tomlinson I.P.M.;  
RA "The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Kudsoon's 'two-hit' hypothesis.;"  
RT Kudsoon's 'two-hit' hypothesis.;"  
RL Med. 5:1071-1075(1995).  
CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE  
CC -1- CATENIN. ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN.  
CC -1- SUBUNIT: FORMS HOMOLOGOMERS AND ASSOCIATES WITH CATENINS.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES TYPES.  
CC -1- PTM: PHOSPHORYLATED BY GSK-3B.  
CC -1- DISEASE: DEFECTS IN APC ARE A CAUSE OF FAMILIAL ADENOMATOUS POLYPOSIS (FAP) AND GARDNER'S SYNDROME (GS), THAT CONTRIBUTE TO TUMOR DEVELOPMENT IN PATIENTS WITH NONINHERITED FORMS OF COLORECTAL CANCER. FAP IS CHARACTERIZED BY ADENOMATOUS POLYPS OF THE COLON AND RECTUM, BUT ALSO OF UPPER GASTROINTESTINAL TRACT (AMPULLARY, DUODENAL AND GASTRIC ADENOMAS). THIS IS A VICTIOUSLY PREMALIGNANT DISEASE WITH ONE OR MORE POLYPS PROGRESSING THROUGH DISPLASIA TO MALIGNANCY IN UNTREATED GENE CARRIERS WITH A MEDIAN AGE AT DIAGNOSIS OF 40 YEARS.  
CC -1- DISEASE: APC MUTATIONS HAVE LED TO SOME INTERESTING OBSERVATIONS. (1) THE GREAT MAJORITY OF THE MUTATIONS FOUND TO DATE WOULD RESULT IN "TRUNCATION" OF THE APC PRODUCT. (2) ALMOST ALL THE MUTATIONS HAVE OCCURRED WITHIN THE FIRST HALF OF THE CODING SEQUENCE, AND SONATIC MUTATIONS IN COLORECTAL TUMORS ARE FURTHER CLUSTERED IN A PARTICULAR REGION (CALLED MCR (MUTATION CLUSTER REGION)). (3) MOST IDENTIFIED POINT MUTATIONS IN THE APC GENE ARE TRANSITIONS FROM CYTOSINE TO OTHER NUCLEOTIDES. (4) THE LOCATION OF GERM-LINE MUTATIONS TENDS TO CORRELATE WITH THE NUMBER OF COLORECTAL POLYPS IN FAP PATIENTS. INACTIVATION OF BOTH ALLELES OF THE APC GENE SEEMS TO BE REQUIRED AS AN EARLY EVENT TO DEVELOP MOST ADENOMAS AND CARCINOMAS IN THE COLON AND RECTUM AS WELL AS SOME OF THOSE IN THE STOMACH.

CC -!- SIMILARITY: CONTAINS 7 ARM REPORTS.  
 CC -!- DATABASE: NAME=APC; NOTE=Information about APC mutations;  
 CC URL="http://perso.curie.fr/~therry/Sousi/APC.html".  
 Query Match 5.9%; Score 103; DB 1; Length 2843;  
 Best Local Similarity 22.3%; Pred. No. 14;  
 Matches 57; Conservative 46; Mismatches 107; Indels 46; Gaps 11;  
 RT Matches 57;  
 RT  
 Qy 79 PELERPKVHHRGKILPFKSHGVAFRELRSPTHGQSOSRKVFDFHSYMQOLESTRNE 138  
 DB 2152 PDEOEEKPTSNKGPRKLKPGKEGSTLETKTSESSKGKGGKVV-YKSLITGVRS---- 205  
 DB 2206 - NSEISGQ-----MKQPLGA-NHPSISRGRMIH1PGRNNSSTSSEVKKGPP 222  
 Qy 139 GEPHLAGDRSKHLKFVIRCLRALKRPRISDLMESPFLN--SGEALSPATAKDE 196  
 DB 2205 - :|: ;:|: ;:|: ;:|: ;:|: ;:|: ;:|: ;:|: ;:|: ;:|: ;:|: ;:|:  
 DB 2295 SSKAPSRSGRSDTPSPRAQQPLSRPIOSPPGRNHSISPPNKLSQLPRTSSPSTA 2354  
 Qy 314 ASQQVNRNRNAKSLI 329  
 Db 2355 ST---KSSGGKMSY 2366  
 RESULT 8  
 ABFL\_HUMAN STANDARD; PRT; 3703 AA.  
 ID ABFL\_HUMAN STANDARD;  
 AC Q15911; Q13719; O5101;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)  
 DE (AT-binding transcription factor 1).  
 OC Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC NEBL\_TAXID=9606;  
 RN [1] SEQUENCE FROM N.A. (ISOFORM A).  
 RP TISSUE=Lung;  
 RX MEDLINE=95070776; Pubmed=7592926;  
 RA Mura Y., Tam T., Ido R., Morinaga T., Miki T., Hashimoto T.,  
 RA Tamaoki T.;  
 RT "Cloning and characterization of an ABFL isoform that expresses in a  
 neuronal differentiation-dependent manner.";  
 RT J. Biol. Chem. 270:26840-26848(1995).  
 RN [2] SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE=Hepatoma;  
 RX MEDLINE=9204333; Pubmed=1719379;  
 RA Morinaga T., Yasuda H., Higashio K., Tamaoki T.;  
 RA "A human alpha-fetoprotein enhancer-binding protein, ABFL, contains  
 four homeodomains and seventeen zinc fingers.";  
 RL Mol. Cell. Biol. 11:6041-6048(1991).  
 RN [3] SEQUENCE OF 1-1190 FROM N.A. (ISOFORM A).  
 RX MEDLINE=950425270; Pubmed=10493829;  
 RA Loftus B.J., Kim U.-J., Sheddock V.P., Kalush F., Brandon R.,  
 RA Fuhmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,  
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,  
 RA Elchler E.E., Harris P.C., Venter J.C., Adams M.D.,  
 RT "Genome duplications and other features in 12 Mb of DNA sequence from  
 human chromosome 16p and 16q.";  
 RL Genomics 60:295-308(1999).  
 RP SEQUENCE OF 1151-3703 FROM N.A.  
 RA KOZLOVICZ A., MOURQUET Y., HOTIC M.;  
 CC Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core  
 sequence of the enhancer element of the AFP gene.  
 CC -!- SUBCELLULAR LOCATION: Nucleus.  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms: A (shown here) and B; are  
 produced by alternative splicing.  
 CC -!- SIMILARITY: CONTAINS 4 HOMEOBOX DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 in the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce>  
 or send an email to license@ib-sib.ch).  
 DR SMART; SMD0355; ZNF\_C2H2; 21.  
 DR SMART; SMD0451; ZNF\_C2H2; 7.  
 DR PROSITE; PS00027; HOMEOBOX; 1; 2.  
 DR PROSITE; PS50028; ZINC\_FINGER\_C2H2\_1; 14.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 9.  
 KW Transcription regulation; Activator; Zinc-finger; Metal-binding;  
 KW DNA-binding; Homeobox; Nuclear protein; Repeat; Polymorphism;  
 KW Alternative splicing.  
 FT ZN\_FING 282 305 C2H2-TYPE.  
 FT ZN\_FING 640 663 C2H2-TYPE.  
 FT ZN\_FING 671 694 C2H2-TYPE.  
 FT ZN\_FING 726 750 C2H2-TYPE.  
 FT ZN\_FING 804 828 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 945 968 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 984 1008 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1040 1064 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1088 1112 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1223 1246 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 1252 1275 C2H2-TYPE.  
 FT ZN\_FING 1360 1385 C2H2-TYPE.  
 FT ZN\_FING 1401 1423 C2H2-TYPE.  
 FT ZN\_FING 1429 1452 C2H2-TYPE.  
 FT ZN\_FING 1454 1479 C2H2-TYPE.  
 FT ZN\_FING 1569 1596 C2H2-TYPE.  
 FT ZN\_FING 1596 1620 C2H2-TYPE.  
 FT ZN\_FING 1983 2006 C2H2-TYPE.  
 FT DNA\_BIND 2145 2204 HOMEOBOX 1.  
 FT DNA\_BIND 2242 2301 HOMEOBOX 2.  
 FT ZN\_FING 2328 2351 C2H2-TYPE (ATYPICAL).  
 FT ZN\_FING 2530 2552 C2H2-TYPE.  
 FT DNA\_BIND 2641 2700 HOMEOBOX 3.  
 FT ZN\_FING 2711 2734 C2H2-TYPE.  
 FT DNA\_BIND 2944 3003 HOMEOBOX 4.  
 FT ZN\_FING 3024 3048 C2H2-TYPE.  
 FT ZN\_FING 3253 3255 C2H2-TYPE.  
 FT DOMAIN 104 107 POLY-PRO.  
 FT DOMAIN 460 489 POLY-GLU.  
 FT DOMAIN 770 784 POLY-ALA.  
 FT DOMAIN 1723 1773 POLY-GLN.  
 FT DOMAIN 1789 1794 POLY-GLN.  
 FT DOMAIN 1852 1857 POLY-GLN.  
 FT DOMAIN 2037 2052 POLY-PRO.  
 FT DOMAIN 3197 3209 POLY-GLN.  
 FT DOMAIN 3210 3214 POLY-PRO.  
 FT DOMAIN 3227 3231 POLY-GLN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

---

CC DOMAIN 3376 3389 POLY-GLN.  
FT DOMAIN 3392 3395 POLY-SLN.  
FT DOMAIN 3507 3527 POLY-SLY.  
FT DOMAIN 3597 3600 POLY-PRO.  
FT DOMAIN 3636 3639 POLY-SER.  
FT VARIANT 1 914 MISSING (IN ISOFORM B).  
A -> V.  
/FTID=VAR\_011694.  
MISSING.

FT VARIANT 3377 3384 /FTID=VAR\_011695.  
G -> GCG.  
>/FTID=VAR\_011696.  
A -> S (IN REF. 3).  
A -> P (IN REF. 3).  
T -> A (IN REF. 3).  
I -> S (IN REF. 3).  
A -> V (IN REF. 3).  
HHRV -> RHIG (IN REF. 3).  
A -> S (IN REF. 3).  
EEIEVDEGEPEETAADEPLAKDQEAGASSQAELKDASP  
VPI (IN REF. 3).  
GVGRSPRGLVLLTSRQLLEFQDVTDPAQPH

FT SEQUENCE 3703 AA; 404468 MW; 0F62AF3/D4DCF856 CRC64;

Query Match 5 %; Score 103; DB 1; Length 3703;  
Best Local Similarity 2B.0%; Pred. No. 20;  
Matches 44; Conservative 20; Mismatches 49; Indels 44; Gaps 8;

QY 171 SIDLM--SPPNLSEALSLSTA--TAKDETQMLIKSARSLGIVMVSRRQ 218  
ID :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 2368 SMDAMEILPTPSSCTPMPMSQASAPKSAANTTASFL--TAEEELATENSKTE 2424

QY 219 EFYLRARRRKFAW-----KPVQISCEMKVMEHTTPMAYRDGSP 262  
ID :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 2425 -----AGDEKPKLAEPASQAPQNQKERUGQRPKEPLQ---QQEQPEQKINTPOQ----KL 2472

QY 263 KNASTPSLPGPNKISPRQVSVPRSSPPRNVSPP 299  
ID :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 2473 QLVSLPLSFQPPQAPQPCPLPQ-SSSPSPSOLSHLP 2508

RESULT 9

WS14 MOUSE STANDARD; PRT; 864 AA.

AC 099M23; 099M20; 099M21; 099M22; 099M29;

DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Williams-Beuren syndrome chromosome region 14 protein homolog (Mr 110 kDa) (Protein name).  
DE Interactor(s).  
GN WSCR14 OR M10.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Muriinae; Mus. OC NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPlicing.  
RX MEDLINE=2115101; PubMed=1120181;

RA Cairo S., Melia G., Urbinatti F., Ballabio A., Raymond A.;  
RT "WSCR14, a gene mapping to the Williams-Beuren syndrome deleted region, is a new member of the Mx transcription factor network.",  
RL Hum. Mol. Genet. 10:617-627 (2001).  
CC -1- FUNCTION: TRANSCRIPTION REPRESSOR. BINDS TO THE CANONICAL AND NON-CANONICAL E BOX SEQUENCES 5'-CACCTG-3' (BY similarity).  
CC -1- SUBUNIT: BINDS DNA AS A HETERODIMER WITH TCF4/Mlx (BY similarity).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY). 5 ISOMORPHES 1/2ETA (SHOWN HERE), 2/THETA, 3/1OTA, 4/KAPPA AND 5/ETA, ARE PRODUCED BY ALTERNATIVE SPlicing.  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHHL) FAMILY OF TRANSCRIPTION FACTORS. BHHL-2TP SUBFAMILY.

CC .....

CC

CC DOMAIN 345 350 POLY-SER.  
FT DNA\_BIND 660 674 BASIC DOMAIN.  
FT DOMAIN 700 714 HELIX-LOOP-HELIX MOTIF (POTENTIAL).  
FT VARSPLIC 715 736 LEUCINE-ZIPPER.  
FT VARSPLIC 58 79 MISSING (IN ISOFORM 2).  
FT VARSPLIC 545 556 AKPEQALEKPTM -> WLVLIVPVSQA (IN ISOFORM 5).  
FT VARSPLIC 557 864 MISSING (IN ISOFORM 5).  
FT VARSPLIC 699 744 VSKATTQKTYILMQLQERAMQEAQOLRDETEELNA INLCQ -> GLIOTRLPLALAGEQSINASESGVHPDRAA GTSYAGGAGAR (IN ISOFORM 3).  
FT VARSPLIC 745 864 MISSING (IN ISOFORM 3).  
FT VARSPLIC 699 714 VSKATTQKTYILM -> LPGLANTEAHIGGARR  
MISSING (IN ISOFORM 4).  
FT VARSPLIC 715 864 MISSING (IN ISOFORM 4).  
SQ SEQUENCE 864 AA; 94874 MW; 7E6AFFB04C71B327 CRC64;

Query Match 5 %; Score 102.5; DB 1; Length 864;  
Best Local Similarity 25.1%; Pred. No. 35;  
Matches 44; Conservative 23; Mismatches 69; Indels 39; Gaps 7;

QY 139 GPEPHLAGDRPSKHLKVLFHCLRLPRISDLIMESPLPNLISGEALSP----- 189  
ID :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 469 GPVVPFSVHPLHGYLVPFRH---FVYQDMQPRCKSSSPGGKASPPTLASATAS 524

QY 190 ---TAKADE--ITOMILKSARSLGIVMVSRRQEFYLARRRRAFWAKPVLSISEM 243  
ID :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 525 PTATRARDNPCLTQLL--FRANKEQALEPPTMPTLRRPESQ-----DINSEI 574

QY 244 KPMVFHTMMAYRDGSPPKNNASTPLSPGKPNISPRQVSV--QRSPPPKNVS 296  
ID :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 575 PRARAEFPPIRTPPRPP-----FGPAPLAPPRLVVKERLSPPSGS 621

RESULT 10

EXTN\_TOBAC STANDARD; PRT; 620 AA.

ID EXTN\_TOBAC AC PI3983;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein). GN HRGPNN3.

OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; Asteridae; eudicots 1; Solanales; Solanaceae; Nicotiana. OC NCBI\_TaxID=4097;  
RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=CV; XANTHI; TISSUE=Leaf;

RX MEDLINE=90128263; PubMed=261209;

RA Keller B., Lamb C.J.,

RT Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation.;

RL Dev. 3: 1639-1646(1989).

CC -!- FUNCTION: HAS A SPECIALIZED STRUCTURAL FUNCTION, POSSIBLY IN THE MECHANICAL PENETRATION OF THE CORTEX AND EPIDERMIS OF THE MAIN ROOT.

CC -!- SUBCELLULAR LOCATION: Extracellular matrix.

CC -!- PTM: EXTENSINS CONTAIN A CHAARACTERISTIC REPEAT OF THE PENTAPEPTIDE SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN GLYCOSYLATED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

---

DR EMBL: X13885; CNA32090.1; -.

DR PIR: S05733; S05733.

KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;

FT Hydroxylation. 1 20 POTENTIAL.

FT SIGNAL 21 620 EXTENSIN.

FT CHAIN 70 73 H-A-P-P.

FT REPEAT 148 151 H-A-P-P.

FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.

FT REPEAT 229 245 1. 2. CONTAINS THE SER-PRO(4) REPEATS.

FT DOMAIN 236 242 2. 2. CONTAINS THE SER-PRO(4) REPEATS.

FT DOMAIN 205 600 3 X APPROXIMATE TANDEM REPEATS.

FT DOMAIN 499 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;

Query Match 5.9%; Score 102; DB 1; Length 620; Best Local Similarity 32.4%; Pred. No. 2.5%; Matches 23; Conservative 5; Mismatches 31; Indels 12; Gaps 1;

Matches 23; Conservative 5; Mismatches 31; Indels 12; Gaps 1;

Oy 245 PVMFHPMPMAXRDSGSPRNASNPSPGKPNISP-----RQVSQPRSSPPP 292

Db 378 PPPSFSPPPTEQSPPPPSPAYSPPLAPPTYSPPPTYSPPPTYAOPPLPPTYSPPP 437

Oy 293 KNNSPPPPAFP 303

Db 438 PATSPPPPPTY 448

RESULT 11

IRS2\_MOUSE STANDARD PRT: 1321 AA.

ID IRS2\_MOUSE STANDARD PRT: 1321 AA.

AC P81122;

DT 15-JUL-1998 (Rel. 36, created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Insulin receptor substrate-2 (IRS-2) (4PS).

GN IRS2.

OS Mus musculus (Mouse);

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL\_SEQUENCE.

RX MEDLINE=9505472; PubMed=7655087;

RA Sun X.-J., Wang L.-M., Zhang Y., Yenush L., Myers M.G. Jr., Glashen E., Lane W.S., Pierce J.H., White M.F.;

RT "Role of IRS-2 in insulin and cytokine signalling.";

RL Nature 377:173-177(1995)

CC -!- FUNCTION: MAY MEDIATE THE CONTROL OF VARIOUS CELLULAR PROCESSES BY INSULIN.

---

CC -!- TISSUE SPECIFICITY: SKELETAL MUSCLE, LUNG, BRAIN, LIVER, KIDNEY,

CC -!- HEART AND SPLEEN.

CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 PTB DOMAIN.

CC HSSP; P35568; IIRS.

DR MED; MGI-10934; Trs2.

DR InterPro; IPR02404; Insulin\_Recep\_S-1.

DR InterPro; IPR001849; PH.

DR Pfam; PF02174; IRS; 1.

DR PRINTS; P000538; INSULINSI.

DR SMART; SMD0235; PH; 1.

DR SMART; SMD0310; PB1; 1.

DR PROSITE; PS50003; PH\_DOMAIN; 1.

FT MOD\_RES 671 671 144 PH DOMAIN 10 300 PTB.

FT MOD\_RES 536 536 536 PHOSPHORYLATION (BY INSR)

FT MOD\_RES 649 649 649 PHOSPHORYLATION (BY INSR)

FT MOD\_RES 1242 1242 1242 PHOSPHORYLATION (BY INSR)

FT MOD\_RES 911 911 911 PHOSPHORYLATION (BY INSR)

FT MOD\_RES 970 970 970 PHOSPHORYLATION (BY INSR)

FT MOD\_RES 1303 1303 1303 PHOSPHORYLATION (BY INSR)

FT DOMAIN 19 28 (BY SIMILARITY).

FT DOMAIN 444 449 449 POLY-ASN.

FT DOMAIN 638 641 641 POLY-SER.

FT DOMAIN 935 939 939 POLY-SER.

SQ SEQUENCE 1321 AA; 136526 MW; 5069CE614960C7 CRC64;

Query Match 5.9%; Score 101.5; DB 1; Length 1321; best local similarity 21.8%; Pred. No. 7; Matches 78; Conservative 43; Mismatches 131; Indels 113; Gaps 18;

Matches 78; Conservative 43; Mismatches 131; Indels 113; Gaps 18;

Oy 27 LSKSCSYHDLIQLNPDYVGVGSPVTITNRMMAEEVLEYGVTCATPNAQPFELERPK 86

Db 396 LSRS---HTLSAG----CGGRPSKVI-LAPAGGALQH----SRSNSMPVAHSPEA 438

Oy 87 VVHRSKSKIPRFESKKG-----VHRELRSPFHGSQS-----RK 120

Db 439 ATSPGS--LSSSSGHGSGSTPLPGCSHPLPHLPHOGDRPSGSASAGSPSPDGFM 496

Oy 121 VFDYHSV--TMOOLESTRNGEPP-----HLAGDRPSKHLKLVFLIRIC 161

Db 497 LDEYGGSPGDLRKFSSHRNTPESTAETPPPARSGSGELGYMSMDRP-----LSIC 548

Oy 162 LKALRPLPRSIDMESTPLNUSGEALSPTAENDETQMLSAARELGMYVSKRQEY 221

Db 549 GRPYR--RVSGDQAQDLDRGKRRKSVTTTPAR---QRQVQPSSASLDYTLRATFS 602

Oy 222 LRRAR-----BRRKFAWKVPLGVSISEMR-----PVMEFHPTMAYR 256

Db 603 GSSGRUCPSPASSPKVAVNPYPEDYGDIEGIGSHKSSSNLGADDGMM----TPGAAL 658

Oy 257 DSGSPPRNASTPSL-PGPKNISPPQVSUPORSS-PPKAVWSPPOPAFVART---ASK 310

Db 659 RSGGPNSKSDDYPMSPTSVSAKQIQLPRLAALPPGAVPAPPSSGVGRTPVNGGG 718

Oy 311 YSAAS 315

Db 719 YKASS 723

RESULT 12

RWL\_DROME



DR EMBL: AF281899; AAC53891.1; -. DR HSSP: P9803; ISUH. DR MGD: MGJ:1891428; Pcdh15. DR InterPro: IPR003126; Cadherin. DR Pfam: PF00028; cadherin\_10. DR PRINTS: PR00205; CADHERIN. DR SMART: SM00112; CA; 1.1. DR PROSITE: PS00233; CADHERIN\_1; 4. DR PROSITE: PS0265; CADHERIN\_2; 11. KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat; Signal; Multigene family; Vision. FT SIGNAL 1 26 POTENTIAL. FT CHAIN 27 1943 PROTOCADHERIN 15. FT DOMAIN 27 1381 EXTRACELLULAR (POTENTIAL). FT TRANSMEM 1382 1402 POTENTIAL. FT DOMAIN 1403 1943 CYTOPLASMIC (POTENTIAL). CADHERIN 1. FT DOMAIN 145 152 CADHERIN 2. CADHERIN 3. FT DOMAIN 153 270 CADHERIN 3. FT DOMAIN 153 270 CADHERIN 4. FT DOMAIN 150 400 CADHERIN 5. FT DOMAIN 151 621 CADHERIN 6. CADHERIN 7. FT DOMAIN 622 722 CADHERIN 7. CADHERIN 8. FT DOMAIN 825 931 CADHERIN 9. FT DOMAIN 932 1040 CADHERIN 10. CADHERIN 11. FT DOMAIN 1150 1264 CADHERIN 11. FT DOMAIN 1437 1448 POLY-PRO. FT DOMAIN 1772 1778 POLY-PRO. FT DOMAIN 1804 1812 POLY-PRO. FT CARBOHYD 57 667 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 102 729 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 206 773 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 424 826 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 564 856 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 667 856 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 729 773 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 826 856 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 1069 1069 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD 1089 1180 N-LINKED (GLCNAC. . .) (POTENTIAL). FT CARBOHYD SEQUENCE 1943 AA; 214816 MW; E3D7DB9F5738652 CRC64; SQ

Query Match 5.8%; Score 100; DB 1; Length 1943; Best Local Similarity 23.7%; Pred. No. 15; Matches 63; Conservative 32; Mismatches 93; Indels 78; Gaps 12;

ID APC\_RNT STANDARD: PRT: 2842 AA. DR P70478; DT 16-OCT-2001 (Rel. 40, created) DT 16-OCT-2001 (Rel. 40, last sequence update) DT 16-OCT-2001 (Rel. 40, last annotation update) DE Adenomatous Polyposis coli protein (APC protein). GN APC. OS Rattus norvegicus (rat). OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. OC NCBI\_TaxID:10116; RN [1] RP SEQUENCE FROM N.A. RC STRAIN=FISCHER 34/N; RX MEDLINE=9516966; PubMed=8563176; RA Toyota M., Ushijima T., Kakuchi H., Watanabe M., Imai K., Yachi A., RA Sugimura T., Nagao M.; RT \*DNA cloning of the rat APC gene and assignment to chromosome 18. " RL Mamm. Genome 6:746-748(1995). RN [2] RP MUTAGENESIS. RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N; RX MEDLINE=9518647; PubMed=7846077; RA Kakuchi H., Watanabe M., Ushijima T., Toyota M., Imai K., Weisburger J.H., Sugimura T., Nagao M.; RT \*Specific 5'-GGGA-3'-->5'-GCA-3' mutation of the Apc gene in rat colon tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine. " RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995); CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE ALLOWS THE DOWN-REGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY). CC -1- SUBUNIT: FORMS HOMOOLIGOMERS AND ASSOCIATES WITH CATENINS (BY SIMILARITY). CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY). CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS. CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC --- DR EMBL: D38629; BAA07609.1; -. DR HSSP: Q02248; 3BCM. DR InterPro: IPR00225; Armadillo. DR SMART: SM00185; ARN; 5. DR PROSITE: PS50176; ARM\_REPEAT. DR KW Anti-oncogene; Phosphorylation; Coiled coil; Repeat. DR DOMAIN 1 728 LEP-RICH. DR DOMAIN 1 62 COILED COIL (POTENTIAL). DR DOMAIN 125 260 COILED COIL (POTENTIAL). DR REPEAT 451 493 ARM 1. DR REPEAT 503 545 ARM 2. DR REPEAT 545 589 ARM 3. DR REPEAT 590 636 ARM 4. DR REPEAT 637 681 ARM 5. DR REPEAT 682 723 ARM 6. DR REPEAT 724 765 ARM 7. DR DOMAIN 739 7831 SER-RICH. DR DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC). DR DOMAIN 1356 1575 ASP/GLU-RICH (ACIDIC). DR DOMAIN 1864 1891 HIGHLY CHARGED. DR DOMAIN 1864 1891 C>R: IN AN IC-INDUCED COLON TUMOR. DR MUTAGEN 523 C>R: IN AN IC-INDUCED COLON TUMOR. SQ SEQUENCE 2842 AA; 310530 MW; 3CBB2E8A34E8847 CRC64;

Query Match 5.8%; Score 100; DB 1; Length 2842; Best Local Similarity 21.9%; Pred. No. 23; Matches 59; Conservative 40; Mismatches 97; Indels 74; Gaps 12;





Db	101 NYGSPVHVHNANYSMPSPDMDPADRROPEQARRPLSVATDNMLEFYKKQDGVMVMD	160	Qy	208 -----SELQMYVSKRQEFLRRRKKFAWPKVLOQISERMKPVMEHTPM	253
Qy	110 S---PTGQSQRSKVFDYHYSYTMQLESTRNEPEPHLAGRSRSHLKVFRHCIRALR	165	Db	610 GTACAGTQPGAQPCAQPSQSPQSPADSPHTLRK-----VSKKLAPIPPKVPGQPG	664
Db	161 TNWVARRGSSAGRKV-----SCAPPNSQPPAPPAAELAALPSP-----	198	Qy	254 AYRDSG---SPPKNAST-PSI-LPGPKNISPRQVSVP-----	286
Qy	167 LPRISDLIMESPLNLSEALSP---TATAKEITOMILKSAAR-----	207	Db	668 KAHADQSAGQLSPVSLSPPPSPSPYGLSYPOGVSLASGOLSPAAPPLASPSVFTLS	724
Db	199 LPEQPLDSSPAAPALSPSGLQGPBRTETTKR - ELSPSAQKGSPGSSCTACGTO	256	Qy	287 RSPPPK---NVSPPPAFARTAS 309	
Qy	208 -----SELQMYVSKRQEFLRRRKKFAWPKVLOQISERMKPVMEHTPM - MAYRDG	259	Db	725 KSRPTPKPRORPLPPOPOPTVNLSAS 751	
Db	257 PGAQPGQPGASPSPSQQPADSPHTLRK-----VSKKLAPIPPKVPGQAMADOSAG	311			
Qy	260 SPK---NSTPSLPGPKNISPRQVSVP-----	293			
Db	312 QFSPVSLSPIPPSTPSPGSLSPQGVSLASGOLSPAAPPLASPSVFTLSKSRRTPPK	371			
Qy	294 --NVSPPPAFARTAS 309				
Db	372 RORPTLPPOPTVNLSAS 390				
<hr/>					
RESULT	2		RESULT	3	
ID	075160	PRELIMINARY;	ID	09VZB8	PRELIMINARY;
AC	075160:	PRT;	AC	09VZB8;	PRT;
DT	01-NOV-1998	(TREMBLrel. 08, Created)	DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)	DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)	DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)
DE	KIAA0672	PROTEIN.	DE	CG4964	PROTEIN.
GN	NCBI_TaxID:9606;		GN	CG4964.	
OS	Homo sapiens (Human).		OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		OC	Bacteria; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
RC	[1]		OC	Ephydriidae; Drosophilidae; Drosophila.	
TISSUE-BRAIN:	SEQUENCE FROM N.A.		OX		
RX			RN	[1]	
RA			RP	SEQUENCE FROM N.A.	
RA			RC	STRAIN=BERKLEY;	
RA			RA	MEDLINE=20196005; PUBMED=10731132;	
RA			RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA			RA	Antanaitis J., Burton R.A., Li P.W., Hoskins R.A., Galle R.F.,	
RA			RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA			RA	Sutton G.G., Woitman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA			RA	Brandon R.C., Rogers Y.-H.C., Blazej R.C., Champé M., Pfeiffer B.D.,	
RA			RA	Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,	
RA			RA	Abril J.F., Agbyanli A., An H.J., Andrews P., Frankkoch C., Baldwin D.,	
RA			RA	Sutton G.G., Woitman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA			RA	Brandon R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA			RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA			RA	Borková D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,	
RA			RA	Burkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,	
RA			RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA			RA	Cherry J.M., Cowley S., Dahake C., Davenport L.B., Davies P.,	
RA			RA	De Pablo J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA			RA	Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA			RA	Durban K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleschmann W.,	
RA			RA	Foster C., Gabril A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA			RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA			RA	Harris N.L., Harvey D., Heinlein J.T., Hernandez J.R., Houck J.,	
RA			RA	Houston K.A., Holland T.J., Wei M.-H., Ibegwam C.,	
RA			RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,	
RA			RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA			RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA			RA	Liu X., Mathei T., McIntosh T.C., McLeod M.P., McPherson D.,	
RA			RA	Merkulov G., Mishina N.V., Mobariv C., Morris J., Moskaleff A.,	
RA			RA	Mount S.M., Moy M., Murphy B., Murphy B., Murzy D.M., Nelson D.L.,	
RA			RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,	
RA			RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Purif V., Reese M.G.,	
RA			RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA			RA	Shue B.C., Sider-Kiamos I., Simons M., Skupski M.P., Smith T.,	
RA			RA	Spiel E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA			RA	Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA			RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,	
RA			RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA			RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA			RA	Zheng X.H., Zhong W.N., Zhou X., Zhu S., Zhu X., Smith H.O.,	
RA			RA	Gibbs R.A., Myers E.W., Rubin G.M., Wenner J.C.,	
RA			RA	"The genome sequence of Drosophila melanogaster.";	
RA			RA	Science 287:2185-2195(2000).	
RA			RA	Science 287:2185-2195(2000).	
DR			DR	EMBL: AE003477; AACF47729.1.;	
DR			DR	HSRP: PS2276; ITAK;	
DR			DR	FlyBase: FBgn0035410; CG34964.	

DR InterPro; IPR003962; FNII\_repeat.  
 DR InterPro; IPR003961; FNIII.  
 DR InterPro; IPR003559; Ig\_MHC.  
 DR InterPro; IPR003005; p\_rich\_extensn.  
 DR Pfam; PF00041; fn3; 2.  
 DR PRINTS; PRO0014; FNTYBIL.  
 DR PRINTS; PRO01217; PRICEXTENSN.  
 DR SMART; SM00060; FM3; 2.  
 DR REPEAT; SM00409; IG; 2.  
 DR SEQUENCE 1381 AA; 153482 MW; 8341F17BB4542745 CRC64;  
 DR Best Local Similarity 7.2%; Score 125.5; DB 5; Length 1381; Gaps 15; Matches 72; Conservative 46; Mismatches 135; Indels 69; Gaps 15;  
 DR QY 30 SCSVMDLILGNP-DIYCGSPITINRMAEVL---EIGYATFYCATPNAQPFLE 83  
 DR 367 STPLHDLIGSEKSYKFRKAENPYGILSEPGESSEBLIFPDPKRGITKPKSATRAGEKD 426  
 DR QY 84 OPKVHTRGSKLIPRFSGHGVREURSPHGSQOSRKVFDYHSYMOOLESTRNNEGEPFH 143  
 DR 427 -PKTAGGHQVPR-----RKTLSPPRQDAST-----GMSPKOSSAKRK-PKPQ 471  
 DR QY 144 LAGDPESKIL-KVIFIRHCLRLPRISIDMESPLNLSGEALSPPATAKDEIQMIL 202  
 DR 472 LIDNEQDTHEMSGTSBALK-----MDVKSPSLNSADSANKPTTDSSNPKNLTL 523  
 DR QY 203 KSAASELGGMYYSKRQEFLYLRRRFAWKPVLOSISEMCKPMEEFTPMAYRDGGSP- 261  
 DR 524 VITTLAPLDKSVR-----SPKARTPATSPKLKENP---KPSGAPK 561  
 DR QY 262 -----P-KNASTPSLPGPKNIS--P-PROVSQPOSSRERPKNSP\_P-PROPEVA---R 306  
 DR 562 DRSPVQPKFQPLTPRMEPDKASPNPKRSLSPPNKRQPLRKSPTPEPIKVTPALLR 621  
 DR DE 307 TASKS-AASQWYQRNGNKS 327  
 DR 622 SAEPVOLGVNQNRRESGQTL 643  
 DR  
 DR RESULT 4  
 DR QSLQ09 ID 09LQ09 PRELIMINARY; PRT; 796 AA.  
 DR AC 09LQ09; DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DR DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DR DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DR DE FT6P17.12 PROTEIN.  
 DR GN FT6P17.12.  
 DR OS Arabidopsis thaliana (Mouse-ear cress).  
 DR OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicots; Rosidae; OC Ericidae; Brassicales; Brassicaceae; Arabidopsis.  
 DR OX NCBI\_TaxId=3702;  
 DR RN [1]  
 DR RP SPOUSE FROM N.A.  
 DR RA Llaca V., Lou A., Young S., Messing J.;  
 DR RT "Retrotansposable elements of Sorghum bicolor.";  
 DR RL Submitted (DBC-1998) to the EMBL/GenBank/DDBJ databases.  
 DR EML; AF11475; ADD2567; J.;  
 DR InterPro; IPR002965; p\_rich\_extensn.  
 DR PRINTS; PRO01217; PRICEXTENSN.  
 DR KW HYPOTHETICAL PROTEIN.  
 DR SEQUENCE 1104 AA; 12594 MW; 59A1FE2CB899C855 CRC64;  
 DR  
 DR Query Match 7.1%; Score 122.5; DB 10; length 1104; Gaps 5; Matches 37; Conservative 17; Mismatches 41; Indels 17; Gaps 5;  
 DR QY 228 RRKFAWKPVLOSISEMPVME--HHTPVMAYRQSGSPKNAST\_PSLCPKNSPPQVSY 284  
 DR 548 RKRKFIILPQRDSSNLSPLSPAHISAAGDNAGSPPTPAKPTLPPEP---PPRSPPP 603  
 DR QY 285 PORS-SPPPKNSPPROPAVARKA-----SKYSAASQYORNRNGNAK 326  
 DR Db 604 PPRSPSPRSPRSPRSKRSAPPAPPSPSKRSARSKVSPSPKGSK 655  
 DR Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EML; AC011000; AACF5808.1; -.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR Pfam; PF00226; DnaJ; 1.  
 DR  
 DR RESULT 6  
 DR ID 070495 PRELIMINARY; PRT; 897 AA.  
 DR AC 070495  
 DR PRINTS; PRO01217; PRICEXTENSN.  
 DR SMART; SM00271; DnaJ; 1.  
 DR P550076; DNAJ2; I.  
 DR SEQUENCE 796 AA; 87577 MW; B2E107AB32DE9986 CRC64;  
 DR Query Match 7.2%; Score 124; DB 10; Length 796; Gaps 11; Matches 68; Conservative 36; Mismatches 133; Indels 52; Gaps 11;  
 DR Db 164 TRQSQSQSPOHHQSQOSQPKHNQNOTVSFRQONQOLP-FOLAHOPHFAQVTOQSOSQORS 222  
 DR QY 121 VDYHYSMTMOLESIRNIGPEHLACD---IPSKHILKLFIRICLRLRURISTDM 175  
 DR Db 223 QFDQQLQEQQQASWROQHQHSSGDKWKEVEELINNATEROVSSRFAEPE 282  
 DR QY 176 ESSLPLNLSGEALSPATAKDEET---QNLIAASEL--GMVVKSOEFYLRRR 228  
 DR Db 233 REPSPSEG---SSHRIHPSTEWTASKPPTPPVSEPVRISELPWOYSEPAROTOL--SSRS 337  
 DR QY 229 RFAFWKPVQLOISBMKPVMEFTPMAYRQSG---SPKNNSTPPLGPKNTSPPROVSY 284  
 DR Db 338 SBAQOLSLIPSVDSS---HASQPTRSQSHAVSKRQPSKPPPMSPQPP--TSN 390  
 DR QY 285 PORSSPPPKNSP-----PROPATVARTASKSAAQVORN 322  
 DR Db 391 PPLSOPPSKPPMSOSSONSKPPVSSQSKPLVSSQSK 438  
 DR  
 DR RESULT 5  
 DR 09XBQ2 ID 09YE02 PRELIMINARY; PRT; 1104 AA.  
 DR AC 09YE02; DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DR DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DR DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DR DE HYPOTHETICAL 125.9 KDA PROTEIN.  
 DR OS Sorghum bicolor (Sorghum) (*Sorghum vulgare*).  
 DR OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; OC Panicoideae; Andropogoneae; Sorghum.  
 DR OX NCBI\_TaxId=4558;  
 DR RN [1]  
 DR RP SPOUSE FROM N.A.  
 DR RA Llaca V., Lou A., Young S., Messing J.;  
 DR RT "Retrotansposable elements of Sorghum bicolor.";  
 DR RL Submitted (DBC-1998) to the EMBL/GenBank/DDBJ databases.  
 DR EML; AF11475; ADD2567; J.;  
 DR InterPro; IPR002965; p\_rich\_extensn.  
 DR PRINTS; PRO01217; PRICEXTENSN.  
 DR KW HYPOTHETICAL PROTEIN.  
 DR SEQUENCE 1104 AA; 12594 MW; 59A1FE2CB899C855 CRC64;

01-AUG-1998 ("REMBREL. 07, Created)  
 DT 01-AUG-1998 ("REMBREL. 07, Last sequence update)  
 DT 01-DEC-2001 ("REMBREL. 19, Last annotation update)  
 DE PLENTY\_OF\_PROJECTS=101.  
 GN SRM1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NEBI\_TAXID=10990;  
 RN [1]  
 RP SUBSEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RA Vaysiere B.M., Camonis J.H.;  
 RL Submitted (MAY1998) to the EMBL/GenBank/DDBJ databases.  
 DR EML1; AF062655; ANC17422.1; .  
 DR MGD; MGI:185003; SIRM1.  
 DR InterPro; IPR002483; PMID:  
 DR Pfam; PF01480; PWT: 1.  
 DR SM00311; PWT: 1.  
 DR SEQUENCE 897 AA; 101166 MW; 62160BBB6772BD10 CRC64;

Query Match 6.9%; Score 120.5; DB 11; Length 897;  
 Best Local Similarity 24.4%; Pred. No. 0.04; Mismatches 103; Indels 95; Gaps 17;  
 Matches 75; Conservative 34; Mismatches 103; Indels 95; Gaps 17;  
 Qy 74 TPNAQPFLEQI-----PKVNH-----RESKI-----LDRFSKH-GVHVR 106  
 Db 387 SPASASPRRHRPSSPATPPKRTRHSPTPOQSNTRSRVSPGTSKVKHKGTEK 446  
 Qy 107 ELSRPT-----HQSOSR-----FVFDYHSVMOLESTRNEGRPHLAGRPS 150  
 Db 447 ESPSPAKPRKVSESEBDGSKMAADSSVQORQRQRNQDSSSDSSSISDEBRPK 506  
 Qy 151 KHLKLVIFIRHCLRALRPRTEIDLMEESPPLNUSGEA----LSPATAKDETTQMLIKS 204  
 Db 507 RSHVK-----NGEVRRRLSTSRSASPSPKRKQET 538  
 Qy 205 AARSELGMVYKSKROFYLRARRRKFAWKPVLP-----GSISEKPKVMEFHMPYMAVRDGSQ 261  
 Db 539 SPRMQMKG----KRMQSPVTKSSRRRKSPPSPARRRSPSPARPPPPPPPPRPPRSPTP 594  
 Qy 262 PKNASTPLSPGPKNISP\*PROVSVP-QR---SSSPPK--NTSPRPOPAFVARVASKSA 314  
 Db 595 PRRRTFPS-PPRRRSPSFRYSPSPPKRTTASPPPP--KRRASPPPP 650  
 Qy 315 SQQVORN 321  
 Db 651 KRRVSHS 657

Query Match 6.9%; Score 118.5; DB 5; Length 1102;  
 Best Local Similarity 25.8%; Pred. No. 0.079; Mismatches 149; Indels 59; Gaps 18;  
 Matches 85; Conservative 36; Mismatches 149; Indels 59; Gaps 18;  
 Qy 36 LILGNPDYVVGSTPYTTINRKAEEVLEYGVTVFVCATPNQPFLERQPKVVRHGSKL 95  
 Db 451 LYRILEDYFGNTAGTCIP-----LEPGQDNLVEYTFPKERIQKURKRIFRG-KVL 507  
 Qy 96 PRFKHCVHVRHLSPR-----THGSQSKRKFVDYHS-----VMMOLESTRNEGRP-EPH 143  
 Db 508 ---RH----TELNDIVYIWARIILNSAPSHTPHSPWYIISTQDKSSKQSSPHSPH 559  
 Qy 144 LAG----DRPSHKLVTRHCLRALRPRTEIDLMEESPPLNUSGEA----LSPATAKDETTQ 199  
 Db 560 TSGGSTPKSPTPITPSPPYFDHSGSRKSLSQENGGL-RPPPIP-MSEEDL--IVGHRSATE 615  
 Qy 200 MILKSAARSELGMVYKSKROFYLRARRRKFAWKPVLP-----GSISEKPKVMEFHMPYMAVRD 257  
 Db 616 PI1-PPREFFLGGSVSSDE--HRVDNTGPPPKFIMESWAVSPTHRHTSROHQKPRD 671  
 Qy 258 SG-----SPPKNASPS-----LPGPKNISPPROVSVPORSSSPPKNVSPPPQ 300  
 Db 672 TGREYNVWISHSPRSPVAPTRPSPAHOVFPNNPNSPPTPLS-PSRVAPLPPISQPEPA 730  
 Qy 301 PAFVARTASKVSAASQQVORNQNAKSLI 329  
 Db 731 HTKVNRRIEQS---QKOPMVGNTNNVE 755

RESULT 7  
 ID 062570 PRELIMINARY; PRT; 1102 AA.  
 AC 062570;  
 DT 062570;  
 DT 01-AUG-1998 ("REMBREL. 07, Created)  
 DT 01-AUG-1998 ("REMBREL. 07, Last sequence update)  
 DT 01-DEC-2001 ("REMBREL. 19, Last annotation update)  
 DE PROTEIN KINASE C-RELATED KINASE (PRKSD).  
 OS Suberites domuncula.  
 OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;  
 OC Hadromedida; Superitidae; Suberites;  
 OC NCBI\_TAXID=55567;  
 RN SEQUENCE FROM N.A.  
 RA Mueller W.E.G.;  
 RL Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SUBSEQUENCE FROM N.A.  
 RX MEDLINE=20033487; PubMed=8798342;  
 RA Kruse M., Gamulin V., Cetkovic H., Pancer Z., Mueller T.M.,  
 RA Mueller W.E.G.; Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 "Molecular evolution of the Metazoon protein kinase C multigene

family";  
 RT J. Mol. Evol. 43:374-393(1996).  
 RL CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 DR HSSP; Y13104; CAY3558.1; .  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR000961; Pkinase\_C.  
 DR InterPro; IPR000861; REM\_repeat.  
 DR InterPro; IPR002290; Ser\_thr\_Pkinase.  
 DR Pfam; PF02185; HK1; 3.  
 DR Pfam; PF00669; Pkinase; 1.  
 DR Pfam; PF0033; Pkinase\_C; 1.  
 DR SMART; SM0074; HR1; 3.  
 DR SMART; SM00220; S\_TK\_C; 1.  
 DR SMART; SM00133; S\_TK\_X; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE\_ATP; 1.  
 DR PROSITE; PS00108; PROTEIN KINASE\_ST; 1.  
 DR ATP-binding; Kinase; Serine/threonine protein kinase; Transferase.  
 KW SEQUENCE 1102 AA; 124286 MW; F4F7AB5248323883 CRC64;

RESULT 8  
 ID 09SK28 PRELIMINARY; PRT; 742 AA.  
 AC 09SK28;  
 DT 01-MAY-2000 ("REMBREL. 13, Created)  
 DT 01-MAY-2000 ("REMBREL. 13, Last sequence update)  
 DT 01-DEC-2001 ("REMBREL. 19, Last annotation update)  
 DE AT2G5050 PROTEIN.  
 GN AT2G5050.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyt; eudicotyledons; core eudicots; Rosidae;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TAXID=7027;  
 RN [1]  
 RP SUBSEQUENCE FROM N.A.  
 RC STRAIN=CV; COLUMBIA;  
 RX MEDLINE=20033487; PubMed=10517197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.D., Barnstead M.E., Reidlyum T.V.,  
 RA Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,

RA Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.,  
 RA Copenhagen, G.P.; Preuss, D.; Niemann, W.C.; White, O.; Eissen, J.A.,  
 RA Salzberg, S.L.; Fraser, C.M.; Venter, J.C.; "Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 RT thaliana.", *Nature* 402:761-768(1999).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV; COLOMBIA;  
 RA Lin X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL: AC006585; AD23008.1; .  
 DR InterPro: IPR002965; P-RICH\_EXTENSN.  
 DR PRINTS; PRO1217; PRICHEXTENSN.  
 SQ SEQUENCE 742 AA; 81758 MW; EBSE7A4096F57D28 CRC64;

Query Match 6.4%; Score 111.5; DB 5; Length 539;  
 Best Local Similarity 21.7%; Pred. No. 0.12; Gaps 6;  
 Matches 43; Conservative 38; Mismatches 75; Indels 31; Gaps 6;  
 Matches 60; Conservative 48; Mismatches 127; Indels 42; Gaps 9;  
 QY 52 TITNRMAAEVLEVGIVYFVCATPNAQPFPLEROPKVVHRGSKILPFRSHGVREURSP 111  
 Db 352 TVFNQQTAAANLQESLSLGSSTRSPRSRSLSESLEYKKRKLW-----ISENTVSSP 404  
 QY 112 THGSGQSKRKYFDHSVNMQLESITRNGEPEPHAGDRSKHLKLFIRCLRALRIPRIS 171  
 Db 405 DTSSEPEK-----DMSSHKSAD-PNSILKKVDESRGLYSVORNVHSKIFSPRMV 456  
 QY 172 IDLMESPQLNSELSEBALSPAKDELTMQILKAANSQELM-----YUSKROEFYLRA 225  
 Db 457 OSPVPSLPLNRSPTQGPASTSRFH-----SSFSLLGITSILHDGSKCDEESTSS 508  
 QY 226 RRRKFPAWKPVLOSIKMKPYNEF---HTPMAYRDSGSPKNASTPSLPGPKNTISPRO 281  
 Db 509 PASPSISFLPFLTQSPLSQPKASQCSQSPTRVHSNCPPEARAVTSSPLP-LKPLRT 567  
 QY 282 VSVPOFSPPPKNVSPPPOPARVARTASYKSAMSQQV 318  
 Db 568 LSRP----PPP----PPPPPISSURSTSPPSSTNSI 596

RESULT 9

Q9TZA0 PRELIMINARY; PRT; 539 AA.

AC Q9TZA0;  
 DT 01-MAY-2000 (Trembleur, 13, Created)  
 DT 01-MAY-2000 (Trembleur, 13, Last sequence update)  
 DT 01-DEC-2001 (Trembleur, 19, Last annotation update)  
 DE HYPOTHETICAL. 62.1 KDA PROTEIN.  
 GN W07B3\_2.  
 OS Cenorhabditis elegans.  
 OC Rhabditidae; Peloderaidae; Caenorhabditis.  
 OC Rhabditida; Pelodera; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV; COLOMBIA;  
 RA Johnson, A.F., la Bastide, M., Lodhi, M., Hoffman, J., Hasegawa, A.,  
 RA Groj, L., Gottesman, T., Granat, S., Hamied, A., Kaplan, N., Schutz, K.,  
 RA Shchedry, N., van Keuren, K., Parment, N., Dedhia, N., Marttienssen, R.,  
 RA McCombie, W.;  
 RA [2]  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV; LANDSBURG;  
 RA T111 S., Grant, S., Parnell, L., Kaplan, N., Hoffman, J., Lodhi, M.,  
 RA Johnson, A.F., Dedhia, N., Marttienssen, R., McCombie, W.R.;  
 RA [3]  
 RP Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA Speigel, L.A., Huang, E.N., Nascimento, L.U., de la Bastide, M., Vil, D.M.,  
 RA Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A., Rodriguez, M.,  
 RA Shekher, M., Schutz, K., See, L.H., Swaby, I., Habermann, K., Dedhia, N.N.,  
 RA Meves, H.W., Lamcke, K., Mayer, K.F.X.;  
 RA [4]  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RA RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL: AF001308; AAC78701.1; .  
 DR EMBL: AF001555; NAB57800.1; .  
 DR EMBL: AL161493; CAB8702.1; .  
 KW Hypothetical protein.  
 SQ SEQUENCE 202 AA; 22315 MW; CAA7EEDE62FAA8AF CRC64;

Y	Query Match	Best Local Similarity	Score 21.9%; Pred No. 0.041;	Length 202;	DB 10;	Score 111; DB 10;	Length 202;		
Matches	54;	Conservative	24;	Mismatches	89;	Indels	80;	Gaps	7;
b	16	RWVYPDGHENVUNISKSCSVHDLNLPNDPYVCGSTP-----YTITNRWA--AEEVL	63	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
b	19	KVVLSDGRVQNLEBETVAEMLNPQHVVVFDPSSISFNNDAKTVKRKLAPDAKTL	78	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
y	64	EVGVIVFYCATPNAQPFLEROPKVYHRGSKILPRSKHGHVHVRERPHQSOSQRKVFD	123	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
b	79	EFGKIVL-----LPA-KRSGRAKSSAVLTSEEVKRL-113	113	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
y	124	YHSVIMQOLESIRNEGPHEPLAGRPRSKHLKLVLFRHCLRALRPRSIDLMESPLNUS	183	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
b	114	-----PSATAVRSSFSYEGILPWF	135	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
y	184	GEALSPATAKDTOMILKSAASELGMVSKROEFYLARRRFAPIWVLOSIQM	243	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
b	136	TRSYKNNPNDT--TVVAAISVGRLAEEMEDRPEFLSRLQLSGR--GWKPSLDPIREK	190	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
y	244	KFVMEFH	250	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
b	191	KAKKKIH	197	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
RESULT	11								
b	919174	PRBLIMINARY;	PRT;	393 AA.					
b	Q9HTI4	PRBLIMINARY;	PRT;	393 AA.					
b	Q9HTI4	"Analysis and Characterization of the Complete Genome of <i>Tupaila</i> (Tree Shrew) Herpesvirus"; J. Virol. 75:4854-4870(2001).	PRT;	393 AA.					
b	01-DEC-2001	(TREMBLrel. 19, last sequence update)	PRT;	393 AA.					
b	01-DEC-2001	(TREMBLrel. 19, last annotation update)	PRT;	393 AA.					
b	T117	<i>Tupaila herpetiviruss</i> . Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae; NCBI_TaxID=10397.	PRT;	393 AA.					
b	[1]	SEQUENCE FROM N.A.	PRT;	393 AA.					
b	AC	STRAIN=2;	PRT;	393 AA.					
b	AC	MEDLINE=21211637; PubMed=11312357;	PRT;	393 AA.					
b	Bahr U., Darai G.; "Analysis and Characterization of the Complete Genome of <i>Tupaila</i> (Tree Shrew) Herpesvirus"; J. Virol. 75:4854-4870(2001).	PRT;	393 AA.						
b	[2]	SEQUENCE FROM N.A.	PRT;	393 AA.					
b	AC	STRAIN=2;	PRT;	393 AA.					
b	Darai G., Bahr U.; Submitted (JUN 2000) to the EMBL/GenBank/DDBJ databases. EMBL: AF281817; ANI57163.1; F4BA27BB3C53BDFF CRC64; SEQUENCE 393 AA; 41784 MW; NCBI_TaxID=10397.	PRT;	393 AA.						
Query	Match	Best Local Similarity	Score 6.3%; Pred No. 0.49;	Length 1073;	DB 10;	Score 109.5; DB 12;	Length 393;		
Matches	89;	Conservative	46;	Mismatches	147;	Indels	95;	Gaps	9;
Db	QY	1 MGLTPFSCV---TQGYVRRVVPDGH---ENISKSCSVHDLNLPNDPYVCGSTP	50	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	507	VGLITTSKVYTMGSTVYQGNNINADKLPCLVEDKLTKDC-VEETCAYHVATLSRN	565	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Qy	51	YTITNRMAEELVEGVTVFVCPMPQELERPKVHKGSKILPRSKHGVREIRS	110	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	566	EVFTWGKGANGRLGHDGDVKDRAFTPLDALKERAVKNAGSNTAACIHL---KWS	620	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Qy	111	PTHGQQS--RKVDFY-----HSYTMQLSIR----NEEPHEPLAGRPS	150	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	621	GTEGNGQACRACERFTRKRHNCNCGLVHCSKSKLSKALAPNG-KPRVCH	677	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Qy	151	KHLKLVFTRHCLRALRPRISOLMESPLNUSGE-----ALSPTATADEIT	198	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	678	CHSKL-----SKVSEANIDSKRVMPLSGENKDRLDKTEIRLAQGIPSNIDLK	728	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Qy	199	QMIKKSARSELGMYVSKROEFYLRARRRKFAWKPIQLSSEMPPMVEHIPMARDS	258	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	729	QDUNRAAROGK-----KADTFSLVRSQTPLTQLDKADLNVADR-----	768	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Qy	259	GSPPKNASTPSLPGPKNISPROQYSVQRSPPRNPVNPQQ-----PAFVARTSKY	312	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	769	RGPPPKPAVTP-----SSSRVPSPRSRSPRSVPVPLNGLGFTSIAESLKTN	820	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Qy	313	-AASQVQRNRNAKS	328	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
Db	821	ELLNQEUVRLRAQESL	837	Q9FHK1	PRELIMINARY;	PRT;	1073 AA.		
RESULT	13								



Search completed: September 16, 2002, 08:51:11  
Job time: 475 sec



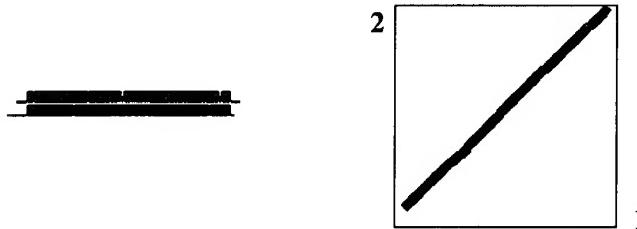
# Blast 2 Sequences results

[PubMed](#)   [Entrez](#)   [BLAST](#)   [OMIM](#)   [Taxonomy](#)   [Structure](#)
**BLAST 2 SEQUENCES RESULTS VERSION BLASTP 2.2.3 [Apr-24-2002]**

Matrix BLOSUM62 ▼ gap open: 11 gap extension: 1  
x\_dropoff: 50 expect: 300. wordsize: 3 Filter  Align

**Sequence 1** gi\_2851522 Mitogen-activated protein kinase 12 (Extracellular signal-regulated kinase 6) (ERK-6) (ERK5)  
(Stress-activated protein kinase-3) (Mitogen-activated protein kinase p38 gamma) (MAP kinase p38 gamma). **Length 367 (1 .. 367)**

**Sequence 2** gi\_1136298 WIPK [Nicotiana tabacum] **Length 375 (1 .. 375)**



NOTE: The statistics (bitscore and expect value) is calculated based on the size of nr database

Score = 292 bits (747), Expect = 5e-78  
Identities = 149/339 (43%), Positives = 219/339 (63%), Gaps = 8/339 (2%)

Query: 21 WEVRAVYRD-LQPVGSGAYGAVCSAVDGRTGAKVAIKKLYRPFQSELFAKRAYRELRLLK 79  
+E+ YR + P+G GAYG VCS ++ VA+KK+ F + AKR RE++LL+  
Sbjct: 36 FEITTKYRPPIMPIGRGAYGIVCSVNLTELNEMVAVKKIANAFDIYMDAKRTLREIKLLR 95

Query: 80 HMRHENVIGLLDVFTPDETLDFTDFYLVMPPFMGTDLGKLMK-HEKLGEDRIQFLVYQML 138  
H+ HENVIGL DV P +F+D Y+ M TDL +++++ ++ L ED Q+ +YQ+L  
Sbjct: 96 HLDHENVIGLIRDVIPP-LRREFSDVYIATELMDTDLHQI IRSNQGLSEDHCQYFMYQLL 154

Query: 139 KGLRYIHAAGIIHRDLKPGNLAVNEDCELKILDGLARQ--ADSEMTGYVVTRWYRAPEV 196  
+GL+YIH+A ++HRDLKP NL VN +C+LKI DFGLAR + MT YVVTRWYRAPE+  
Sbjct: 155 RGLKYIHSANVLHRDLKPSNLLVNANCDLKICDFGLARPNIENENMTEYVVTRWYRAPEL 214

Query: 197 ILNWMRYTQTVDIWSVGCIMAEMITGKTLFKGSDHLDQLKEIMKVGTGTPPAEFVQLQSD 256  
+LN YT +D+WSVGCI E++ K LF G DH+ Q++ + ++ GTP + LQ++  
Sbjct: 215 LLNSSDYTAIDVWSVGCIFMELMNRKPLFGGKDHVHQIRLLTELLGTPTEADLGFLQNE 274

Query: 257 EAKNYMKGLPELEKKDFASILTNASPLAVNLLEKMLVLDAEQRVTAGEALAHPYFESLHD 316  
+AK Y++ LP+ ++ A + + +PLA++L++KML D +R+T EAL HPY LHD  
Sbjct: 275 DAKRYIRQLPQHPRQQLAEVFPHVNPLAIDLVDKMLTFDPTRRITVEALDH PYLA KLHD 334

Query: 317 TEDEPQVQKYDDSFDDVDRTL--DEWKRVTYKEVLSFKP 353  
DEP + SFD + + ++ K + Y+E LS P  
Sbjct: 335 AGDEP-ICPVPFSFDFEQQGIGEEQIKDMIYQEALSLNP 372

CPU time: 0.08 user secs. 0.01 sys. secs 0.09 total secs.



Lambda K H  
0.320 0.137 0.405

Gapped  
Lambda K H  
0.267 0.0410 0.140

Matrix: BLOSUM62  
Gap Penalties: Existence: 11, Extension: 1  
Number of Hits to DB: 1022  
Number of Sequences: 0  
Number of extensions: 86  
Number of successful extensions: 4  
Number of sequences better than 300.0: 1  
Number of HSP's better than 300.0 without gapping: 1  
Number of HSP's successfully gapped in prelim test: 0  
Number of HSP's that attempted gapping in prelim test: 0  
Number of HSP's gapped (non-prelim): 1  
length of query: 367  
length of database: 330,426,180  
effective HSP length: 123  
effective length of query: 244  
effective length of database: 219,683,991  
effective search space: 53602893804  
effective search space used: 53602893804  
T: 9  
A: 40  
X1: 16 ( 7.4 bits)  
X2: 129 (49.7 bits)  
X3: 129 (49.7 bits)  
S1: 41 (21.8 bits)  
S2: 60 (27.7 bits)

